Theory and data for area summation of contrast with and without uncertainty: Evidence for a noisy energy model

Tim S. Meese

School of Life and Health Sciences, Aston University, Birmingham, United Kingdom

🏠 ✉

Robert J. Summers

School of Life and Health Sciences, Aston University, Birmingham, United Kingdom

🏠 ✉

Contrast sensitivity improves with the area of a sine-wave grating, but why? Here we assess this phenomenon against contemporary models involving spatial summation, probability summation, uncertainty, and stochastic noise. Using a two-interval forced-choice procedure we measured contrast sensitivity for circular patches of sine-wave gratings with various diameters that were blocked or interleaved across trials to produce low and high extrinsic uncertainty, respectively. Summation curves were steep initially, becoming shallower thereafter. For the smaller stimuli, sensitivity was slightly worse for the interleaved design than for the blocked design. Neither area nor blocking affected the slope of the psychometric function. We derived model predictions for noisy mechanisms and extrinsic uncertainty that was either low or high. The contrast transducer was either linear ($c^{1.0}$) or nonlinear ($c^{2.0}$), and pooling was either linear or a MAX operation. There was either no intrinsic uncertainty, or it was fixed or proportional to stimulus size. Of these 10 canonical models, only the nonlinear transducer with linear pooling (the noisy energy model) described the main forms of the data for both experimental designs. We also show how a cross-correlator can be modified to fit our results and provide a contemporary presentation of the relation between summation and the slope of the psychometric function.

Introduction

Contrast sensitivity improves with the area of a sine-wave grating. For gratings presented in the center of the visual field, the summation function has a characteristic bowed shape (pressing in towards the origin) when plotted on double log coordinates of threshold versus area (Robson & Graham, 1981; Tootle & Berkley, 1983; García-Pérez, 1988; Rovamo, Luntinen, & Nasanen, 1993; Rovamo, Mustonen, & Nasanen, 1994; Foley, Varadharajan, Koh, & Farias, 2007; Meese & Summers, 2007). It is thought that the initial improvement over small areas owes to spatial filtering (e.g., Meese & Summers, 2007; see supplementary material in Meese, 2010) and that the asymptotic effect at much greater stimulus diameters owes to retinal inhomogeneity (Howell & Hess, 1978; Robson & Graham, 1981; Foley et al., 2007; Meese & Summers, 2007). The intermediate region has a log-log slope of approximately −1/4 (Meese, Hess, & Williams, 2005; Meese & Summers, 2007) and for this reason is sometimes called fourth-root summation (Bonneh & Sagi, 1999). However, its interpretation is controversial, as we describe below.

Probability summation and Minkowski summation

One interpretation of fourth-root summation is in terms of probability summation between multiple independent mechanisms, each responding to different regions of the stimulus. According to one approach, if the contrast transducer is linear and followed by a high threshold and there are negligible false-positive responses, then probability summation predicts a (log-log) summation slope equal to $-1/\beta$, where β is the Weibull slope parameter of the psychometric function (Quick, 1974; Watson, 1979; Robson & Graham, 1981). In this model, β depends on the distribution of internal additive noise placed before the threshold (Sachs, Nachmias, & Robson, 1971; Quick, 1974; Tyler & Chen, 2000; Mortensen, 2002), which is sometimes assumed to be Weibull (Quick, 1974; Graham, 1989). Predictions for probability summation can be derived by setting $\gamma = \beta$ in a generalization of Minkowski summation over m detecting mechanisms as follows:

$$resp_{overall} = \left[\sum_{i=1:m} (|r_i|^{\gamma'}) \right]^{1/\gamma} \tag{1}$$

where, for the conventional implementation, $\gamma' = \gamma$. This

gives the desired summation slope of $-1/\beta$ (Quick, 1974). Empirical estimates of the slope of the psychometric function (β) at detection threshold are $\beta \approx \gamma' \approx 4$ in area summation studies, consistent with the (high-threshold) probability summation model (Robson & Graham, 1981; Meese & Williams, 2000) and the fourth-root empirical description (Meese et al., 2005). Some studies have also found $\beta \approx \gamma'$ for summation of superimposed components that differ in orientation and/or spatial frequency (Sachs et al., 1971; Meese & Williams, 2000), though other studies have found marked differences between β and γ' for related stimulus arrangements (Meinhardt, 2000; Manahilov & Simpson, 2001). The reason for these differences is not clear.

Minkowski summation has been used widely in contrast detection studies where it has enjoyed much success, though it is important to realize that its equivalence to probability summation (Quick, 1974) is rooted in high-threshold theory. Note that high-threshold theory is also the basis for any approach that treats psychometric functions as "probability of *detecting* functions" and then combines the probabilities from multiple "detectors" using conventional statistical procedures (e.g., Graham, Robson, & Nachmias, 1978). By implication, or otherwise, this approach assumes that visual *detectors* can enter a *state* that indicates they have correctly detected the stimulus—and that is high threshold-theory, of course. Unfortunately though, high-threshold theory has been roundly rejected. For example, contrast detection thresholds depend on guess-rate even after correction for guessing, which is inconsistent with the theory's predictions (Green & Swets, 1966; Nachmias, 1981). Nevertheless, the demise of the theoretical underpinning for Minkowski summation as an implementation of probability summation (where $\gamma = \gamma' = \beta$) has not deterred investigators from using it as a method of combining mechanism outputs, and several defenses of this position have been made (Wilson, 1980; Nachmias, 1981; Meinhardt, 2000; Tyler & Chen, 2000; Mortensen, 2002). Indeed, models of early spatial vision tend to remain rooted in the idea that an array of independent filter-elements is followed by a nonlinear pooling strategy and a decision variable (Wilson & Bergen, 1979; Rohaly, Ahumada, & Watson, 1997; Tyler & Chen, 2000; Párraga, Troscianko, & Tolhurst, 2005). As already mentioned, this is usually interpreted as probability summation and implemented using Minkowski summation with exponents $\gamma' = \gamma \approx 3$ or 4. Nonetheless, there is no direct evidence to support the probability summation interpretation, merely the consistency between psychophysical summation data and model predictions (see also the discussion in Robson & Graham, 1981, and Mortensen, 1988). Therefore, we refer to the association between probability summation and area summation of contrast as the first dogma of spatial vision (Meese & Baker, 2011).[1]

Signal combination and Minkowski summation

Graham (1989) and others have emphasized an alternative interpretation of Minkowski summation in terms of "deterministic nonlinear pooling." Thus, the pooling strategy in multiple filter models might be reinterpreted as neuronal convergence (Graham, 1989) or, as we prefer to say, "signal combination." It is rarely presented that way though (but see Watson & Ahumada, 2005), presumably because the indiscriminate arrangement of inputs with which it is often associated seems an unlikely neuronal wiring scheme. However, recent psychophysical work involving area summation of narrow-band stimuli has edged towards this deterministic interpretation (Graham & Sutter, 1998; Meese et al., 2005; Foley et al., 2007; Meese & Summers, 2007, 2009; see also Laming, 1988; Manahilov & Simpson, 1999; Meinhardt, 2000). Specifically, Meese and Summers (2007) suggested that several processes are involved in area summation of grating contrast as follows: retinal inhomogeneity, spatial filtering, contrast transduction ($\approx C_i^p$) at each location (i), the addition of independent (Gaussian) additive noise (G_i) at each location, linear spatial summation across filter-elements, and finally a decision variable. Note that some authors (e.g., Graham, 1989) describe such arrangements as involving *nonlinear summation* because each signal line is subject to a nonlinear transformation of signal contrast before summing. Typically, however, we prefer the term *linear summation* with respect to this situation, by reference to the linearity of the pooling process. This distinguishes it from the nonlinear MAX operator, which is often treated as the "minimal combination rule" (Tyler & Chen, 2000) and is involved in contemporary treatments of probability summation (Pelli, 1985) (see the section on the MAX operator). In previous work we have referred to the preceding model arrangement (Meese & Summers, 2007) as the transducer and noise combination model or more simply, the combination model (Meese, 2010). Here, we introduce the term *noisy energy model* to describe the same arrangement (though not necessarily including retinal inhomogeneity and filtering).

The noisy energy model involves a cascade of two different processes, each of which produces too much summation when operating in isolation (though see Meese & Hess, 2007; Manahilov, Simpson, & McCulloch, 2001), but when combined produce the desired fourth-root effect. For example, a square-law contrast transducer ($p = 2$) followed by summation and late additive noise predicts that detection thresholds decline with a (log-log) slope of $-1/2$ when plotted against stimulus area. This is essentially the contrast energy model (Rashbass, 1970; Manahilov & Simpson, 1999) with late noise. Similarly, when the noise is placed early (before summation) and the transducer is linear ($p = 1$),

we have the ideal summation model (Campbell & Green, 1965; Tyler & Chen, 2000), where area summation of signal and noise also causes thresholds to decline with a (log-log) slope of $-1/2$. This is because the total signal strength is proportional to area and the total noise is proportional to the square-root of area (the noise variances add and the standard deviation is equal to the square-root of their sum). Thus, the signal to noise ratio increases in proportion to area/$\sqrt{}$(area) and the reciprocal of this gives the relation for contrast sensitivity: a power law with an exponent of $-1/2$. In both cases then, the effect of area on contrast sensitivity at a fixed criterion level of performance can be described using Minkowski summation with $\gamma' = 2$ (quadratic summation). But when these two effects are cascaded, as in the noisy energy model—where noise comes after contrast transduction but before summation—then the model predicts a fourth-root summation rule, effectively $\gamma' = 4$ (Wilson, 1980; Meese & Summers, 2007, 2009; Meese, 2010).

Thus, although very different in architecture, the probability summation model and the noisy energy model each make a fourth-root prediction for area summation of contrast.

The slope of the psychometric function and Minkowski summation

Two major benefits of using Minkowski summation to combine the outputs of visual mechanisms are its flexibility and computational simplicity. However, we advise that it be used with caution, as we now explain. Meese and Summers (2009) assumed that early sensory noise was additive and used stimuli designed to encourage spatial pooling over a fixed retinal extent. The aim was to hold the internal noise level constant at the decision variable while allowing contrast area to be manipulated within this window. If successful, this would allow a clean empirical measure of the (nonlinear) transducer without the potentially confounding effects from variably pooled internal noise. Meese and Summers (2009) devised a specific test of a generalized version of Minkowski summation where they set $\gamma' = p\gamma$ in Equation 1 (see also Manahilov & Simpson, 1999). With this model arrangement, p controls the slope of the psychometric function and $p\gamma$ controls the level of summation. They found the best model predictions with $\gamma = 1$ and $2 \le p \le 3$. When $p = 2$, this is equivalent to the noisy energy model. Meese and Summers (2009) showed that the conventional use of Minkowski summation in their model ($\gamma \approx 3$ or 4 and $p = 1$) completely failed to predict the slope of the psychometric function. Thus, while Minkowski summation might offer a pragmatic solution to the problem of combining the outputs of multiple mechanisms, if the model is to make successful predic-

tions for a range of performance levels (i.e., $resp_{overall}$ is not a constant in Equation 1) (Bird, Henning, & Wichmann, 2002; Meese, Georgeson, & Baker, 2006; García-Pérez & Alcalá-Quintana, 2007) then the generalized version of Minkowski summation (Equation 1, where $\gamma' = p\gamma$) should be used, at the very least.

The MAX operator and probability summation

From the discussion above it is clear that (a) models of probability summation associated with high-threshold theory are inadequate, (b) the conventional use of Minkowski summation as an explicit image processing stage is flawed if models are required to make predictions at more than one performance level (i.e., different levels of percent correct or d'), and (c) a model involving linear summation of contrast following nonlinear contrast transduction (the noisy energy model) is consistent with area summation results. However, we must also consider the contemporary formulation of probability summation, which involves the observer taking the MAX over multiple (noisy) mechanisms (Pelli, 1985; Tyler & Chen, 2000). For several situations this model also predicts fourth-root summation slopes ($\gamma' = \gamma \approx 4$; Tyler & Chen, 2000), consistent with area summation data.

In psychophysical experiments, the MAX operator is usually treated as an operation performed by the decision maker, placing it late in the information-processing stream (Pelli, 1985). However, the MAX operator is also a valuable method for computing various image invariances (Riesenhuber & Poggio, 1999), suggesting that there might be sensory (cellular) implementations of the MAX operator arranged throughout the object recognition system (Riesenhuber & Poggio, 2002; Cadieu et al., 2007). On the other hand, the alternating layers of MAX and linear summing operations rising up the hierarchy in these models (Serre, Woilf, Bileschi, Riesenhuber, & Poggio, 2007; Cadieu, Kouh, Pasupathy, Conner, Riesenhuber, & Poggio, 2009) and evidence for both types of summation in areas 17 and 18 of cat (Finn & Ferster, 2007) bring into question any expectation that area summation experiments *should* reveal the MAX operation in visual psychophysics.

Observer uncertainty and the motivation for the present study

Observer uncertainty is used to describe the situation where the MAX operation is applied to a greater number of noisy signal lines than actually contain the target. The level of uncertainty is often summarized as the ratio of irrelevant to relevant mechanisms (U). Uncertainty is an

inherent part of theoretical analyses involving summation and the MAX operator and can affect both the predicted levels of summation and the slope of the psychometric function (Pelli, 1985; Tyler & Chen, 2000; Neri, 2010).

In most previous models of contrast detection, area summation is controlled by the severity of the nonlinear transducer (p) when pooling is linear (Meese & Summers, 2007) or the level of uncertainty (U) when pooling is a MAX operation (Tyler & Chen, 2000). Thus, most previous studies have not been able to distinguish between these two operations because a single parameter can be adjusted to produce similar contrast summation behavior by each of them. We overcame this problem here by including two further factors to help constrain the models. First, we performed the experiment using both interleaved and blocked designs for the various conditions of stimulus area. These designs have different implications for the level of extrinsic uncertainty and consequently the behavior of the models (as we describe later). Second, along with measures of contrast sensitivity, we also analyzed the slope of the psychometric function, which depends on both uncertainty and the form of the contrast transducer (Pelli, 1985; Tyler & Chen, 2000). We then derived signatures for each of our 10 canonical model configurations (2 transducers × 2 pooling rules × 2.5 forms of intrinsic uncertainty) for each of our experimental designs (blocked and interleaved). (Note that the factor of 2.5 derives from the use of three forms of uncertainty for one pooling rule but only two for the other.) A comparison of our analyses and data revealed a single model configuration that produced the correct qualitative relationships between stimulus area and both (a) sensitivity and (b) the slope of the psychometric function for the stimuli used here. The successful model involves a nonlinear contrast transducer followed by additive noise and linear pooling (i.e., the noisy energy model). We then (a) found that successful quantitative predictions could be achieved when retinal attenuation and spatial filtering were included in the model and (b) showed how a conventional cross-correlator can be modified to achieve similar results. Our analyses do not support any of the probability summation (MAX) models.

Methods

Equipment

Stimuli were displayed from the framestore of a Cambridge Research Systems (CRS, UK) ViSaGe stimulus generator operating in pseudo 15-bit mode and were controlled by a PC. The monitor was a NokiaMultigraph 445X (Nokia, Finland) with mean luminance of 76 cd/m^2 and a frame rate of 120 Hz. Look up tables were used to perform gamma correction to ensure linearity over the full range of stimulus contrasts. Observers sat in a dark room at a viewing distance of 74 cm with their head in a chin and headrest and viewed the stimuli binocularly.

Stimuli

Stimuli were always horizontal sine-wave gratings in sine-phase with the center of the display. The spatial frequency was 2.5 c/deg and the stimulus duration was 100 ms. The gratings were modulated by a circular raised cosine function with a central plateau, the diameter of which was the nominal diameter of the stimulus (1–32 cycles). The blurred part of the modulator extended beyond the central plateau and was always 2 pixels wide (12.5 arcmin). All stimuli were sampled with 12 pixels per cycle. Thus, the smallest stimulus had a full diameter of 16 pixels (12 for the central plateau, plus 2 on each side for the blurred boundary) and a diameter at the half height of the envelope of 14 pixels. The largest stimulus had a full diameter of 388 pixels and a diameter at half height of 386 pixels (see Appendix E for stimulus equations).

Fixation marks were displayed throughout the experiment and arranged in one of two ways. For the "fixed quad" marks, a set of four dark square points (12.5 arcmin wide) were displayed at the corners of a virtual square whose virtual contours just surrounded the largest stimulus in the experiment (i.e., the side of the virtual square was equal to the diameter of the largest stimulus). For the "variable quad" marks, the arrangement was similar, but the virtual square was matched to the diameter of each grating. The first arrangement was used in both the blocked design and the interleaved design, whereas the second arrangement was used only in the blocked design. In other words, the blocked design was run twice, with different configurations of fixation marks, whereas the interleaved design was run only once. We avoided using a central fixation point as this can interfere with the detection of small stimuli (Meese & Hess, 2007; Summers & Meese, 2009).

Grating contrast is expressed as Michelson contrast in % [i.e., $c = 100(L_{max} - L_{min})/(L_{max} + L_{min})$] or in dB re 1% [$= 20.\log_{10}(c)$], where L is luminance.

Procedure

Test contrast was selected randomly from seven stimulus levels. The levels were separated by 2.5 dB and centered at approximately each observer's threshold.

There were 140 trials in total (20 trials per stimulus level).

A temporal two-interval forced-choice (2IFC) technique was used, where the stimulus was displayed with a contrast of 0% in one interval (the null interval), and at the contrast selected by the staircase in the other interval (the target interval). The onset of each 100-ms interval was indicated by an auditory tone and the duration between the two intervals was 400 ms. The observer's task was to select the target interval using one of two buttons to indicate his or her response. Correctness of response was provided by auditory feedback, and the computer selected the order of the intervals randomly. For each experimental run, data from the test-stages (above), thresholds (81.6% correct in the absence of lapsing), and the slopes of the psychometric function (β) were estimated by fitting a Weibull function using the psignifit routine (Wichmann & Hill, 2001) where the lapse rate parameter (λ) was a free parameter but constrained such that $0 \leq \lambda \leq 0.05$.

The experiment was performed in three different ways. In an interleaved design, fixed quad fixation was used (see above), and the stimulus size was selected randomly on a trial-by-trial basis. In a blocked design, a sequence of trials was performed to allow an estimate of threshold to be made for a particular size before a subsequent stimulus size was tested. This design was run using the fixed quad fixation and variable quad fixation methods above. In the blocked designs, the order of stimulus size was selected randomly. All stimulus sizes were tested for all three experimental conditions before beginning fresh random sequences. This was done six times to produce estimates of psychometric functions based on 840 trials each.

Observers

The two authors (TSM & RJS) and a postgraduate student (MS) served as observers. The two authors were highly practiced with the task and the conditions. The third observer was naïve to the purposes of the experiment but was experienced with psychophysical procedures. All observers wore their normal optical correction.

Results, Part I: A theoretical study and toy models

It is well known that the shape of the contrast transducer and the level of uncertainty (the proportion of irrelevant mechanisms monitored) can affect levels of summation and the slope of the psychometric function (Pelli, 1985; Tyler & Chen, 2000). Here we extend this work by using Monte Carlo simulations to provide a systematic study of linear and nonlinear (accelerating) contrast transducers with linear or MAX pooling. (Note that our terminology permits the summation process to be linear even if the input to that process has passed through a nonlinear transducer.) Most importantly, we also consider the effects in the context of intrinsic and extrinsic uncertainty. This is controlled by either a free parameter (intrinsic uncertainty, which determines whether the model observer monitors irrelevant noisy channels) or the experimental design (extrinsic uncertainty). Although the analysis below owes much to earlier work by Pelli (1985) and Tyler and Chen (2000), it is the first exposition with sufficient breadth to be able to address the question of the form of summation in our detection experiment. This analysis also forms the basis of Appendix A, where we provide a contemporary account of the relation between the slope of the psychometric function and summation. This supersedes earlier accounts in terms of the Minkowski metric that were based on high-threshold theory (Quick, 1974; Robson & Graham, 1981).

In this section we report toy models in the absence of spatial filtering and retinal inhomogeneity[2] to provide a direct illustration of the effects with which we are concerned. Although both of these factors will be important for a quantitative account of our results (we consider these details when we develop the model in Part IV), there are several qualitative aspects of the models with which they do not interfere (e.g., the ordinal relation of the functions and whether there are design or area effects). These will be the focus of our attention in Part III, following the modeling here. Furthermore, at this stage we are not concerned with the luminance modulation of our stimulus over space and the modulation of responses that this would produce across linear mechanisms (see also Tyler & Chen, 2000). Although this will be picked up in Part IV and Appendix C, these details are largely irrelevant to the empirical study here where the number of sine-wave stimulus cycles was the independent variable. Thus, in the study here, the processing details for each stimulus cycle are of little importance—what matters is the rules that control pooling across multiple cycles.

Canonical summation models

We assumed an array of $m = 1,024$ mechanisms where the number of mechanisms stimulated (s) was equal to the square of the diameter of the stimulus in cycles. Our smallest and largest stimuli excited 1 and 1,024 mechanisms, respectively. A proportional scaling of these figures changed the quantitative details of some of the summation functions (e.g., see Tyler & Chen,
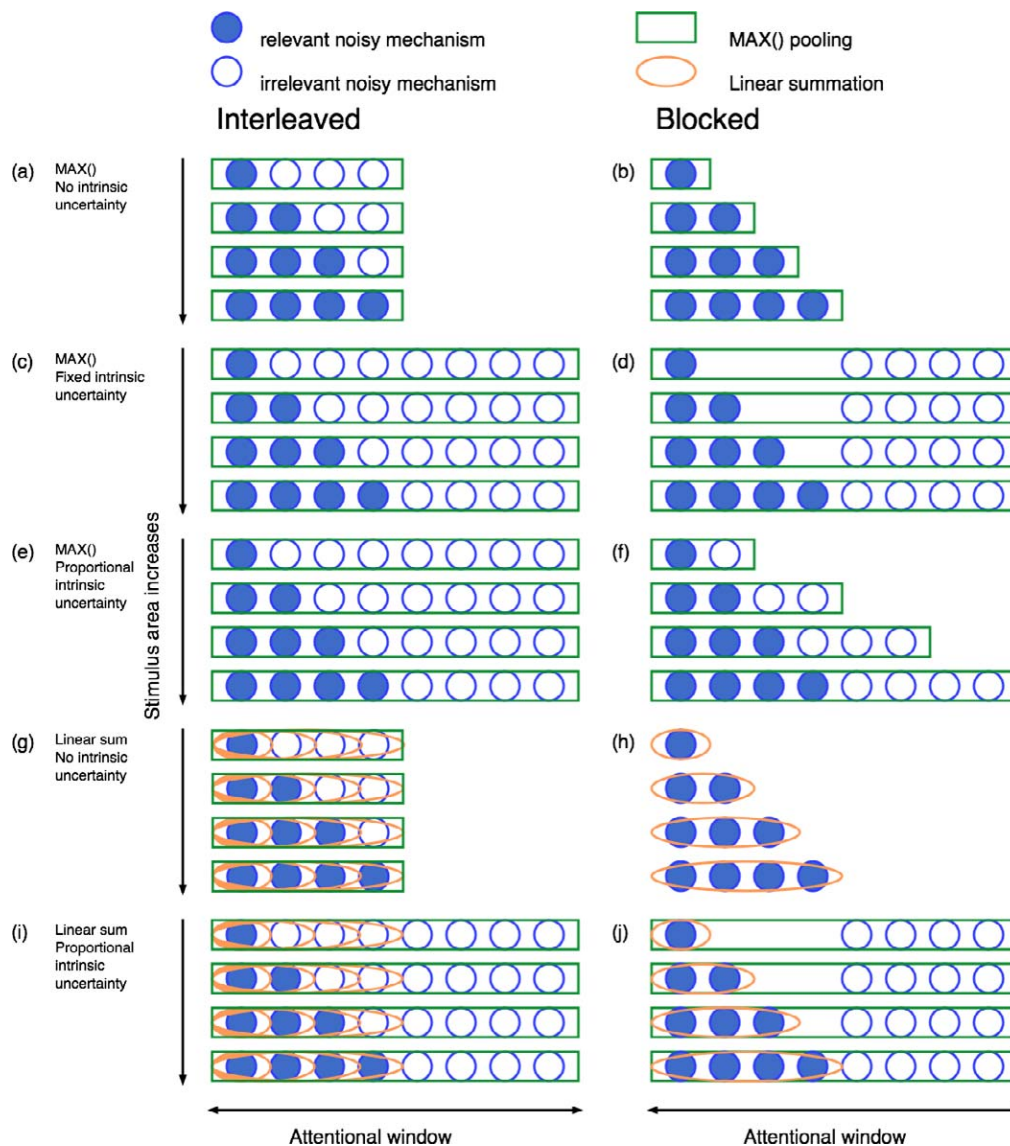
Figure 1. Schematic illustration of the canonical models of spatial summation tested in this paper. Columns are for interleaved and blocked experimental designs. The contrast transducer (not shown) was either linear ($p = 1$) or nonlinear ($p = 2$) giving two times the five different model configurations depicted. The models used in the simulations contained many more mechanisms than those shown here. The eagle-eyed reader might be perturbed that the schemes in (c) and (e) are identical, yet the corresponding red model curves in Figure 2 (c) and (e) are slightly different. This is because of the different number of irrelevant mechanisms involved in the two sets of simulations.

2000, p. 3133; Appendix C), but not their general form or our conclusions. We performed the simulations for $t = 6$ stimulus sizes sequenced in powers of 2 (i.e., 1, 2, 4, 8, 16, and 32).

If a mechanism was stimulated, then the level of excitation (r) was given by the stimulus contrast (c), otherwise it was zero. Every mechanism (i) was subject to contrast transduction followed by independent additive Gaussian noise. The contrast transducer was either linear:

$$f(r_i) = r_i + G_i, \qquad (2)$$

or nonlinear:

$$f(r_i) = r_i^p + G_i, \qquad (3)$$

where an accelerating transducer exponent of $p = 2.0$ was used to conform to the energy model. (For convenience, we sometimes refer to the linear transducer as having $p = 1$.) The parameter G was zero mean, unit variance, Gaussian noise.

In the blocked designs (Figure 1, right) we assumed there was no extrinsic stimulus uncertainty. In the real experiments, when variable quad fixation was used there was a continuous cue to stimulus size from the

Model Parameter	Explanation
r	Mechanism response to stimulus contrast (before noise or nonlinearity)
G	Zero mean, unit variance, additive Gaussian noise (stochastic)
p	Exponent of nonlinear transducer (typically, $p = 2.0$, or $p = 1.0$ for the linear transducer)
m	Number of basic contrast detecting mechanisms that are relevant to the task on at least some of the trials ($m = 1{,}024$)
n	Number of irrelevant noisy mechanisms
s	Number of basic contrast detecting mechanisms excited by the stimulus. This also indicates the relative areas of the stimuli.
t	Number of different stimulus sizes ($t = 6$)
i	Index into the array of $m + n$ basic contrast detecting mechanisms
λ	Number of linear pooling mechanisms (typically, $\lambda = 6$)
j	Index into the array of λ linear pooling mechanisms
$\check{s}_j$	Number of basic contrast detecting mechanisms summed by the jth linear pooling mechanism
σ_j	Standard deviation of the response of the jth linear pooling mechanism
$resp$	Decision variable
$resp_j$ and $resp'$	Intermediate stages in calculating the decision variable.

Table 1. Summary of parameters used in the models. Numbers in parentheses indicate parameter values used in the main simulations, where appropriate.

fixation marks throughout, so this was reasonable. In the fixed quad design observers were soon able to judge the size of the condition from their successful trials, but it is possible that there was some residual extrinsic uncertainty associated with this condition. In sum, for our simulated blocked designs and in the absence of intrinsic uncertainty, the model observer monitored only the set of mechanisms that was excited by the stimulus.

For the interleaved experimental design (Figure 1, left) the observer could not know which mechanism or set of mechanisms were the most appropriate to monitor on each trial. Thus, although details vary across models (see below) every mechanism in the model contributed to the model observer's decision. In general, this observer was extrinsically uncertain.

A major distinction between two classes of model is whether pooling over first-stage (basic) contrast mechanisms is linear (Σ) (sometimes referred to as signal combination) (Figure 1, bottom two rows) or probabilistic, according to a MAX operator (sometimes referred to as probability summation or signal selection) (Figure 1, top three rows). We consider both of these here. Note that when the first-stage pooling was linear (orange ellipses in Figure 1) the outputs of the various mechanisms were normalized to have the same expected variance. This was important when there was uncertainty about which linear pooling mechanism was most appropriate (e.g., the interleaved design) and second-stage MAX pooling was used to choose between them (see Figure 1).

In further simulations we also included MAX pooling over additional noisy mechanisms to model intrinsic uncertainty. For "fixed" intrinsic uncertainty, the observer always monitored an additional fixed set of irrelevant mechanisms regardless of the details of the experimental design or the stimulus condition (Figures 1c, d, i, and j). In the simulations the number of irrelevant mechanisms was set to $n = 1{,}024$. For "proportional" intrinsic uncertainty, the observer monitored additional mechanisms that were a fixed multiple of what would otherwise be monitored. For example, this could happen in a blocked area summation experiment if the observer was certain about position and area in each condition, but always uncertain about spatial frequency and orientation. In this case, as the number of relevant mechanisms increases with area, the number of irrelevant mechanisms also increases in direct proportion. In the main simulations we used a factor of 100. Proportional intrinsic uncertainty was implemented (and relevant) only for MAX pooling of the first-stage mechanisms.

Appropriate combination of the various situations above gave a total of 10 canonical models for each of our two experimental designs (Figure 1). We provide mathematical descriptions of each of these situations, where the two or three forms of intrinsic uncertainty (none, fixed and proportional) are described within each subsection (Table 1 provides easy reference to parameters).

In general, we say that the level of uncertainty (U) is the ratio of the number of mechanisms that are equally excited by the stimulus to the number of mechanisms that are monitored. Thus, when there is no uncertainty, $U = 1$. For completeness, we give explicit expressions for the levels of extrinsic (U_{ext}), intrinsic (U_{int}), and total (U_{tot}) uncertainty below and summarize these in Table 2. These offer some insight into model behaviors since it is well known that for a linear transducer, the effects of uncertainty on the psychometric function are

Pooling method and experimental design	Extrinsic uncertainty U_{ext}	Intrinsic uncertainty U_{int}	Total uncertainty U_{tot}
MAX, interleaved	m/s	$n+1$	$(m+n)/s$
MAX, blocked	1	$(n+s)/s$	$(n+s)/s$
Σ, interleaved	K	$n+1$	$K+n$
Σ, blocked	1	$n+1$	$n+1$

Table 2. Summary of uncertainty for two different pooling methods and two different experimental designs. The model parameters (m, n, and s) are summarized in Table 1. The parameter K is an unknown constant, >1. These expressions were not an explicit part of the computational models, which used stochastic noise and Monte Carlo simulations.

approximately proportional to $\log(U_{tot})$ (Green & Swets, 1966; Pelli, 1985).[3] However, these expressions were not used in generating our model predictions, which relied on Monte Carlo simulations and did not require explicit formulations for uncertainty.

MAX, interleaved (3 models $\times$ 2 transducers)

The model response for the interleaved design with MAX pooling was given by:

$$resp = \underset{i=1:m+n}{MAX}[f(r_i)], \tag{4}$$

where m is the number of mechanisms stimulated by the largest stimulus and n is the number of additional irrelevant mechanisms that control intrinsic uncertainty. In the main simulations, $m = 1{,}024$ (Figure 1a). Intrinsic uncertainty was set according to $n = 0$ (none), $n = 1{,}024$ (fixed), or $n = 99m$ (proportional). For these models, $U_{ext} = m/s$, $U_{int} = n+1$ and $U_{tot} = (m+n)/s$, where s is the number of mechanisms excited by each stimulus. Note that in this instance, the fixed ($n = 1{,}024$) and proportional ($n = 99m$) intrinsic uncertainty models have the same form (m was a constant); they differed only in the overall level of uncertainty (Figures 1c and e).

MAX, blocked (3 models $\times$ 2 transducers)

The model response for the blocked design with MAX pooling was given by:

$$resp = \underset{i=1:s+n}{MAX}[f(r_i)], \tag{5}$$

where s is the number of mechanisms excited by the stimulus in the block (Figure 1b). The parameter n controlled intrinsic uncertainty as follows: $n = 0$ (none), $n = 1{,}024$ (fixed), or $n = 99s$ (proportional). For these models, $U_{ext} = 1$, $U_{int} = (n+s)/s$ and $U_{tot} = (n+s)/s$. Note that here, the level of total uncertainty (U_{tot}) did

not vary with the size of the stimulus (s) when intrinsic uncertainty was proportional to s. On the other hand, U_{tot} decreased with s when intrinsic uncertainty was fixed. This means that these two models of intrinsic uncertainty (Figures 1d and f) have distinct forms for the blocked design.

$\sum$, interleaved (2 models $\times$ 2 transducers)

For the linear pooling we assumed a set of pooling mechanisms for λ different pool sizes, evenly spaced in a logarithmic sequence. For the main simulations here, $\lambda = 6$, and the pooling mechanisms were matched to the $t = 6$ stimulus sizes. However, this was not critical: we found that setting $\lambda = 11$ by adding intermediate pooling mechanisms between the $t = 6$ stimulus sizes had a negligible effect on model behavior.

The number of mechanisms pooled by the jth pooling mechanism is given by $\check{s}_j$. The response of each of these $j = 1{:}\lambda$ linear pooling mechanisms was given by

$$resp_j = \sum_{i=1:\check{s}_j}[f(r_i)]. \tag{6}$$

In the interleaved design the observer could not know the size of the stimulus (s) on each trial, and so a MAX operator was used to select the most responsive of the λ linear pooling mechanisms.[4] However, the expected level of response will increase with j in the absence of stimulation because of the linear pooling of noise. To combat this bias, the response of each linear pooling mechanism was normalized by the expected standard deviation (σ_j):

$$resp' = \underset{j=1:\lambda}{MAX}\left[\frac{resp_j}{\sigma_j}\right], \tag{7}$$

where $\sigma_j = \sqrt{\check{s}_j}$ (Tyler & Chen, 2000). When there was no intrinsic uncertainty (i.e., $n = 0$); $resp = resp'$ (Figure 1g). Otherwise, we had

$$resp = \underset{i=m+1:m+n}{MAX}[resp', f(r_i)], \tag{8}$$

where $n = 1{,}024$ for fixed intrinsic uncertainty (Figure 1i).

Unfortunately, there was a general problem here in deriving an expression for extrinsic uncertainty. This was because each linear pooling mechanism received a graded level of total excitation, depending on the area of the stimulus. This meant that uncertainty could not be expressed simply as the ratio of mechanisms equally excited to those monitored (as we have defined it). Nevertheless, we observed that the distribution of signal-to-noise ratios within the set of linear pooling mechanisms was fairly (though not quite) constant with stimulus area. Therefore, we assumed an equivalent

extrinsic uncertainty approximated by the unknown constant K, where $K > 1$. Thus, for the models here we had $U_{ext} \approx K$, $U_{int} = n + 1$, and $U_{tot} \approx K + n$. Recall that the contents of Table 2, including this approximation, were not part of our formal analysis, which used Monte Carlo simulations (see *Monte Carlo simulations* section).

$\sum$, blocked (2 models $\times$ 2 transducers)

For linear summation in the blocked design, the model observer knew the size of the stimulus (s). In the main simulations there was always a pooling mechanism j, where $\check{s}_j = s$, so we had:

$$resp' = \sum_{i=1:s}[f(r_i)]. \tag{9}$$

As in the interleaved design, when there was no intrinsic uncertainty (i.e., $n = 0$), $resp = resp'$, which is the ideal observer (Figure 1h). Otherwise:

$$resp = \underset{i=m+1:m+n}{MAX}\left[\frac{resp'}{\sqrt{s}}, f(r_i)\right], \tag{10}$$

where $n = 1,024$ for fixed intrinsic uncertainty (Figure 1j). For the models here, $U_{ext} = 1$, $U_{int} = n + 1$, and $U_{tot} = n + 1$.

Monte Carlo simulations

To derive model behaviors, a simulated method of constant stimuli was used with contrasts spaced in 2-dB steps. In a single simulated trial, Gaussian noise was drawn independently for each mechanism in each 2IFC interval. The stimulus was presented in just one interval and the simulated observer used each of the model equations to calculate a response for each interval. On each trial the simulated observer selected the interval containing the largest response and the trial was marked as correct if it contained the target. Five thousand trials were simulated at each contrast level and a threshold (α; 81.6% correct) and slope of the psychometric function (β) were estimated by fitting a Weibull function to the simulated data. The guess rate for the psychometric function was set to 50%, appropriate for 2IFC, giving:

$$\hat{\Psi}(c) = 0.5 + 0.5\left\{1 - \exp\left[-(c/\alpha)^\beta\right]\right\}. \tag{11}$$

Confidence limits (95%) for these stochastic models were calculated using a bootstrap technique. These are denoted by the pale regions around the model curves in the figures, where large enough to be seen.

An alternative approach would have been to derive analytic expressions for each of our models. In some cases (e.g., the linear pooling models with no uncertainty) this is very straightforward (e.g., Meese & Summers, 2007; Appendix D), but in others, less so. For example, when the summation rule is the MAX operator, the problem remains tractable but becomes more complicated (e.g., see Tyler & Chen, 2000). Therefore, in the interests of simplicity and transparency (of exposition), we chose to use Monte Carlo simulations throughout.

Toy model behaviors

Figures 1 and 2 show toy model detection thresholds for probability summation (the MAX operator) and linear summation, respectively. They are functions of the number of stimulated mechanisms (s, equivalent to stimulus area) for the linear and nonlinear transducers
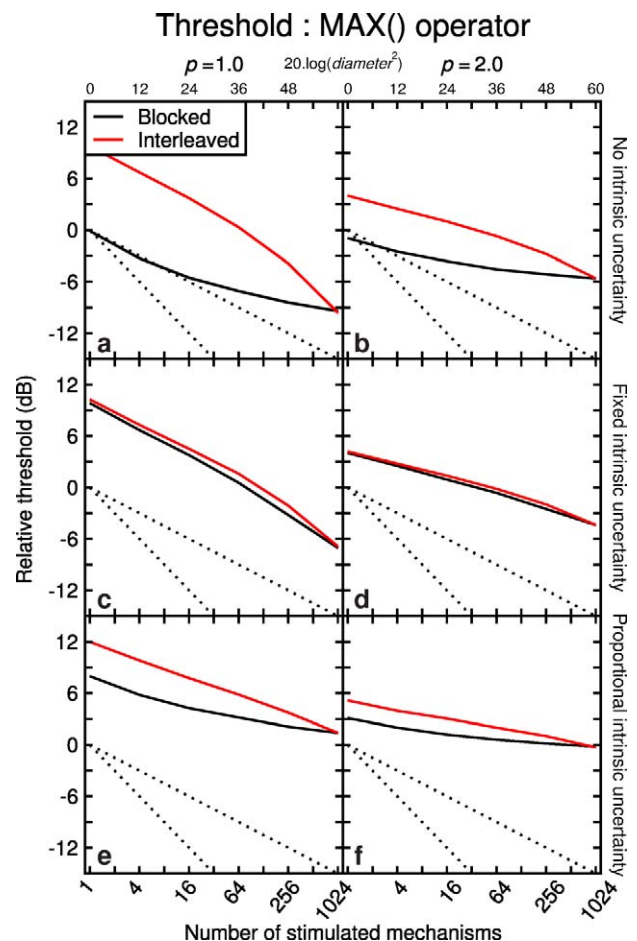

Figure 2. Summation slopes (i.e., contrast thresholds as functions of area) for MAX pooling, two transducer exponents (different columns) and three different forms of intrinsic uncertainty. Extrinsic uncertainty was set by the experimental design, which was either blocked (black curve) or interleaved (red curve). The dotted lines have slopes of $-1/4$ and $-1/2$ for comparison. Note the double log axes.

(left and right columns). Within each panel, the pair of curves is for the interleaved (red) and blocked (black) designs. The intrinsic uncertainty is, from top to bottom: none, fixed, and proportional (proportional is for the MAX operator only). The dotted lines show summation slopes of $-1/4$ (fourth-root summation) and $-1/2$ (quadratic summation) for comparison.

Contrast sensitivity and summation

Here we describe some of the key effects on contrast sensitivity in the models and provide intuitive explanations where appropriate. Readers who are not interested in these details could skip to Part II of the results without loss of continuity.

For the linear transducer and the MAX operator with no intrinsic uncertainty (Figure 2a), the summation curves (for blocked and interleaved designs) each have a slope close to $-1/4$ for the initial part of the function, though they deviate from this as the number of mechanisms stimulated increases (see Tyler & Chen, 2000 for discussion). For most stimulus sizes, performance is much better for the blocked design than the interleaved design, owing to the absence of extrinsic uncertainty in the blocked design. The main effect of adding fixed intrinsic uncertainty (Figure 2c) is to reduce the distance between these two functions. Although there is drop in model performance for both designs, this is most substantial for the blocked design, which brings the two functions closer together. In essence, intrinsic uncertainty makes a substantial contribution to the expression for total uncertainty (see Table 2), thereby reducing the impact of extrinsic uncertainty in the interleaved design. When the intrinsic uncertainty is proportional to the number of mechanisms otherwise monitored (Figure 2e), some separation remains between the functions for blocked and interleaved designs, but sensitivity is reduced quite markedly for the larger stimulus sizes. This results in rather shallow summation slopes.

For linear summation (Figure 3a) and the blocked design (black curve) the model sits on the quadratic summation slope, confirming the well-known result that for the ideal observer, the signal to noise ratio improves with the square-root of the number of mechanisms stimulated. Overall performance for the interleaved design is slightly worse than for the blocked design owing to the extrinsic uncertainty over the size of the pooling mechanism (Equations 6 and 7). Note that the extrinsic uncertainty only reduces overall sensitivity, but does not change the form of the function. This is because the summation slope reveals the operation within the linear pooling mechanisms, and the level of uncertainty does not affect this. Not surprisingly, adding intrinsic uncertainty to the blocked
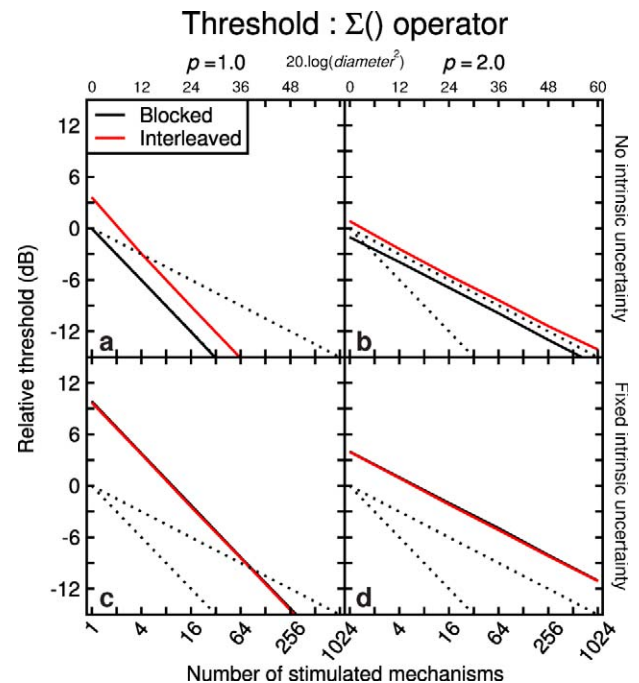

Figure 3. Similar to Figure 2 but for pooling by linear summation and only two forms of intrinsic uncertainty (different rows).

design (Figure 3c) has a similar effect: it decreases overall sensitivity, but does not change the summation slope (compare Figures 3a and c).

For all five model-configurations (different rows in Figures 2 and 3), the effect of replacing the linear transducer with an accelerating transducer (right columns) is to decrease the gradient of the summation slope (e.g., see Meese [2010] for an explanation). It also decreases the size of the design effect, if it were present. This is because a nonlinear transducer acts in a very similar (though not identical) way to uncertainty (Pelli, 1985), and therefore dilutes the extrinsic uncertainty inherent in the interleaved design.

The slope of the psychometric function

The slopes of the psychometric functions for MAX and linear pooling are shown in Figures 4 and 5, respectively. For the linear transducer, MAX pooling, no intrinsic uncertainty (Figure 4a), and the interleaved design (red curve), the psychometric slope decreases with area, confirming the analysis of Tyler and Chen (2000). For the blocked design, the psychometric slope remains at $\beta \approx 1.3$ (equivalent to a d' psychometric slope of unity) across the entire stimulus range. This is to be expected from a linear transducer when there is no stimulus uncertainty ($\beta \approx 1.3$ is the signature of a linear system). The effect of introducing intrinsic uncertainty (Figure 4c) is to increase the slope of the psychometric function (e.g., Pelli, 1985). This has the
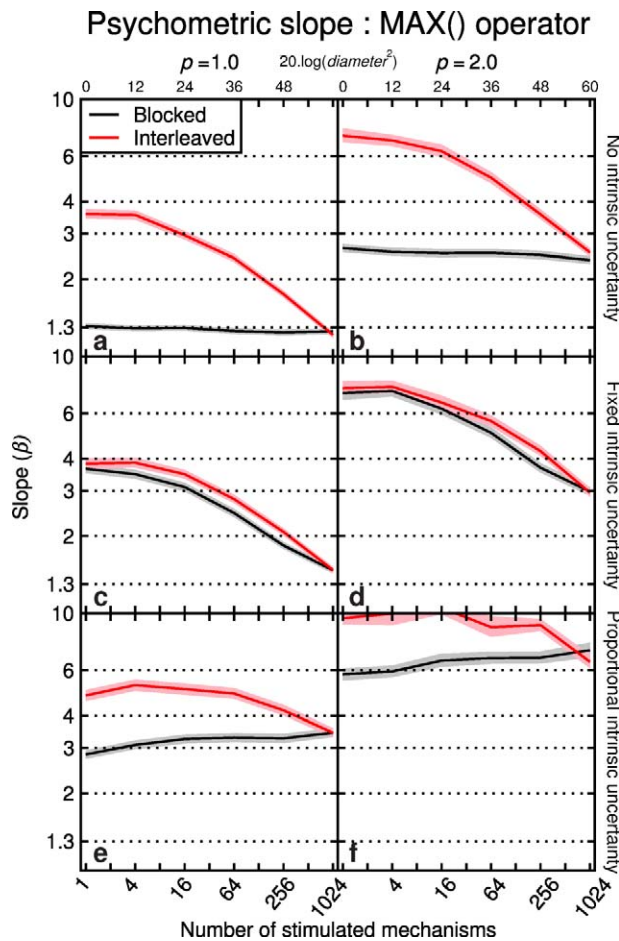
Figure 4. Slope of the psychometric function as a function of the number of mechanisms stimulated (e.g., stimulus size) for MAX pooling, two transducer exponents (different columns), and three different forms of intrinsic uncertainty. Extrinsic uncertainty was set by the experimental design, which was either blocked (black curve) or interleaved (red curve). Note the double log axes.

greatest effect for the small stimulus sizes where the addition of irrelevant noisy mechanisms most seriously compromises the overall signal to noise ratio. This has a large effect in the blocked design (where previously there was no uncertainty), but little effect in the interleaved design, where the log of total uncertainty (the crucial measure) is increased only marginally by the extra mechanisms. The consequence is that the psychometric slopes for the two designs are fairly similar. When the intrinsic uncertainty is proportional to the number of mechanisms otherwise monitored (Figure 4e) the design effect and its interaction with stimulus size remains intact and the psychometric slopes are steeper overall (typically, $\beta > 3$).

When both pooling and the transducer are linear (Figure 5a) and the design is blocked, the psychometric slope is $\beta \approx 1.3$ because there is no uncertainty. It is slightly steeper in the interleaved design because of the
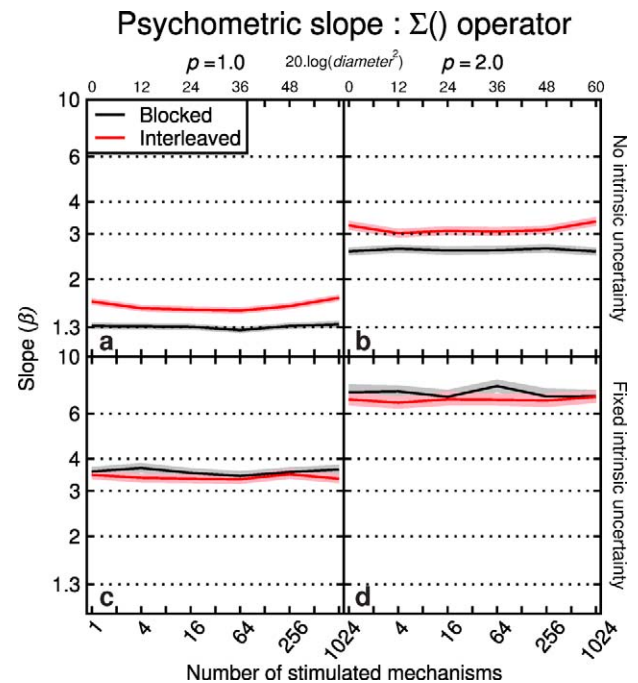

Figure 5. Similar to Figure 4 but for pooling by linear summation and only two forms of intrinsic uncertainty (different rows).

low level of extrinsic uncertainty over the size of the pooling mechanism (Equation 6).

The main effect of replacing the linear transducer with an accelerating transducer (different columns) is to make all of the psychometric functions steeper, for both linear and MAX pooling (Figures 4 and 5). This is because nonlinear contrast transduction and uncertainty have similar effects on the slope of the psychometric function and their effects combine (i.e., the slopes in the right hand columns of Figures 4 and 5 are shifted vertically from those in the left-hand columns). Thus, the relation between the slopes expected for the two different designs is unaffected by the choice of transducer.

Results, Part II: An empirical study

Area summation for sine-wave gratings: Summation slopes

The results of the psychophysical experiment are shown in Figure 6 averaged across the three observers. The summation functions (Figure 6a) have a familiar bowed form (Robson & Graham, 1981; Rovamo et al., 1993; Foley et al., 2007; Meese & Summers, 2007) where performance improves quite steeply at first, but subsequently more gently. For all three observers there
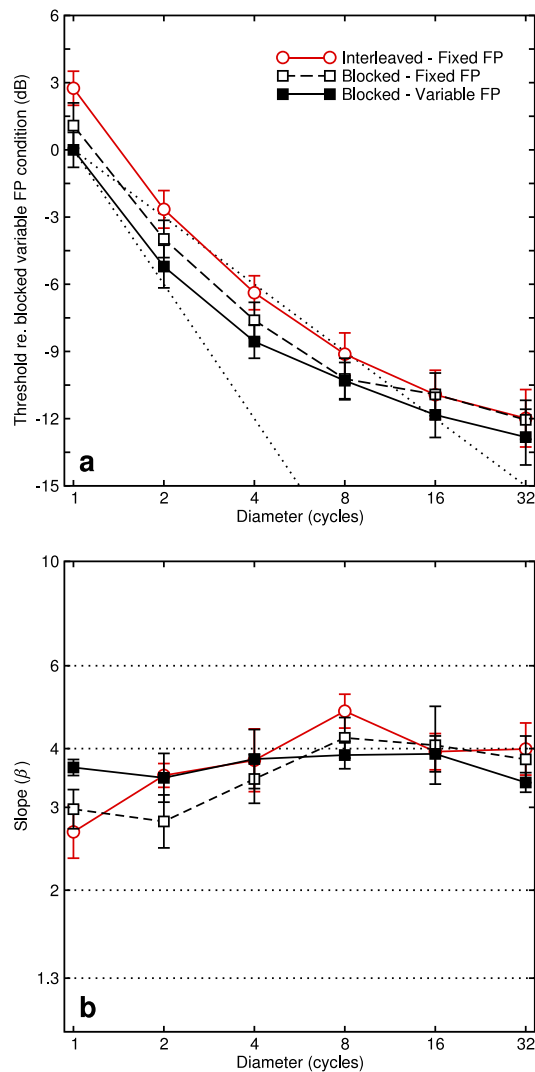
Figure 6. Results from the area summation experiment averaged across three observers. (a) Normalized thresholds as functions of stimulus area for each of three experimental designs (see legend). The average absolute threshold for the interleaved condition was 8.1 dB (re 1%). The dotted lines have slopes of $-1/4$ and $-1/2$ for comparison. (b) Slopes of the psychometric functions as functions of stimulus area for the same three experimental designs. Note that the x-axis here and in later figures refers to the integer number of cycles across the central plateau of the stimulus. Error bars show ± 1 SE across observers. FP: fixation points.

were significant effects of experimental design (interleaved, fixed quad blocked, variable quad blocked) and stimulus area. For TSM there was also a significant interaction between these two factors. For MS this interaction approached but did not reach significance (see Table 3 for statistical details).

On average, performance in the blocked variable quad-fixation design (filled squares and continuous black curve) was better (≈ 3 dB) than in the interleaved fixed quad design (open circles and continuous red curve) at the smaller stimulus sizes. For the large stimulus sizes, the differences were a little smaller. The results for the blocked fixed quad-fixation design (open squares and dashed black curve) tended to be intermediate to the other two.

Meese et al. (2005) also compared the detection thresholds for blocked and interleaved designs, but for a narrower range of stimulus sizes and for a spatial frequency of 1 c/deg. A very small effect was found (though overlooked) in that study (≈ 0.6 dB) for two out of three observers, and was in the same direction as that found here. In that study, the effect might have been less evident because (a) the range of stimulus sizes was less and (b) the central fixation point used in that study did not provide a cue to stimulus size, as the variable quad did here. These factors would decrease the design effect because (a) extrinsic uncertainty would be less and (b) there would be a diluting effect of higher intrinsic uncertainty. Consistent with this hypothesis, the design effect here was smaller for fixed quad fixation than variable quad fixation (compare the open and closed black squares in Figure 6a).

Foley et al. (2007) also compared detection thresholds across blocked and interleaved designs but found no systematic effect for the average of their two observers. We wondered whether this was because the design effects are less easily revealed when only two different stimulus sizes are used, as in the Foley et al. study (one stimulus was 16 times larger than the other). For example, with that arrangement, it is possible that the observer monitors only the outputs of two different-sized pooling mechanisms (a small one and a large one) in which case, the extrinsic uncertainty might have been low, and/or hidden by intrinsic uncertainty, common to each design. To test this general idea we ran our model with spatial filtering and retinal inhomogeneity installed (see Part IV) and for just two different

	df		MS		RJS		TSM	
Source		Error	F ratio	p	F ratio	p	F ratio	p
Design	2	10	6.109	0.018*	59.101	<0.001*	13.304	0.002*
Area	5	25	350.541	<0.001*	494.070	<0.001*	333.792	<0.001*
Interaction	25	50	3.222	0.077	0.561	0.838	2.465	0.017*

Table 3. Two-factor ANOVA for the threshold results for each observer. Asterisks indicate significant effects.

	df		MS		RJS		TSM	
Source	Source	Error	*F* ratio	*p*	*F* ratio	*p*	*F* ratio	*p*
Design	2	10	1.522	0.265	1.303	0.314	0.959	0.416
Area	5	25	0.497	0.775	3.259	0.021*	2.989	0.030*
Interaction	25	50	1.029	0.433	0.812	0.619	0.492	0.888

Table 4. Two-factor ANOVA for the slopes of the psychometric functions for each observer. Asterisks indicate significant effects.

sized pooling mechanisms (matched to stimulus diameters of 1 and 16 cycles, as in the Foley et al. experiment). With this arrangement (and no intrinsic uncertainty), the model predicted a design effect of −0.8 dB for blocked relative to interleaved. This is the size of the effect found for one of the Foley et al. observers (VHN = −0.83 dB) but in the wrong direction for the other (JMF = 0.25 dB). However, owing to the small size of these effects it is difficult to reach a firm conclusion.

Area summation for sine-wave gratings: Slopes of the psychometric functions

The slopes of the psychometric functions are shown in Figure 6b. The geometric means of the slopes were $\beta = 3.67$, $\beta = 3.51$, and $\beta = 3.72$ for the blocked variable, blocked fixed, and interleaved designs, respectively. By eye, there was no systematic variation of psychometric slope with stimulus size or experimental design in the average plot (Figure 6b). However, two-way ANOVA revealed a significant effect of stimulus size for RJS and TSM (see Table 4). Inspection of the data suggested that this was due to upward trends over the first parts of the functions in the interleaved condition and the blocked fixed fixation point condition for RJS and TSM, respectively. However, one-way ANOVA on each of the data sets from each design condition (i.e., nine analyses on three functions for each of three observers) found no significant effects. More importantly, however, we found no evidence for the decrease in the slope of the psychometric function with stimulus area that was predicted by several of the MAX models (see Figure 4).

Results, Part III: Qualitative comparisons between toy models and data

Of the two blocked designs, the one most likely to reduce intrinsic uncertainty, and thereby reveal a design effect of extrinsic uncertainty, was the variable quad design. This was because the fixation marks provided a consistent cue to stimulus size. Therefore, we compared the results from this and the interleaved design (the solid black and red curves in Figure 6) with the various model predictions described earlier (Table 5). Specifically, we were looking for models that produced design and area effects on contrast sensitivity, but no effects on the slope of the psychometric function (β), which should be around 3 or 4. Further simulations (e.g., see Part IV) confirmed that spatial filtering and retinal inhomogeneity had little effect on the slopes of the psychometric functions but caused the summation functions to bow in a similar way to the experimental data. For simplicity, our toy models did not include these processes and so they are not expected to predict the bowing of the summation slopes. Therefore, we overlook mismatches in the third data column of Table 5 for now. For similar reasons, we do not consider whether the models produce an interaction between area and design on the summation slopes. However, all other gross qualitative mismatches between model and data lead to model rejection and are represented by an X in Table 5. Following this procedure all but one of the models was rejected by our data, though several entries are worthy of further consideration.

Near(ish) misses

In Figure 2a the predicted design effect for MAX pooling was much larger than that found in the experiment. The size of the effect can be reduced in the model by decreasing the level of extrinsic uncertainty, which is achieved by reducing the number of mechanisms that are involved in the detection process (i.e., decreasing m). However, the utter failure of this model to predict the slope of the psychometric function (Figure 4a) cannot be remedied.

Another way to decrease the design effect on the summation curves in the MAX pooling model is to add a fixed level of intrinsic uncertainty (Figure 2c), but this also predicts that the psychometric slope should decrease with stimulus area for both designs (Figure 4c), which is inconsistent with the results (Figure 6b). This problem can be overcome in the model by increasing the level of intrinsic uncertainty further (i.e., increasing n). However, this further decreases the design effect on contrast sensitivity (Figure 2c), which

	Contrast sensitivity			Slope of the psychometric function				Reject
	Area effect	Design effect	Sum. slope	Area effect	Design effect	Interleaved $\sim\beta$	Blocked $\sim\beta$	
Human result	**Yes**	**Yes**	**Mid**	**No**	**No**	**3 → 4**	**3 → 4**	**—**
MAX, LT No IU	Yes	Yes	Mid	Yes X	Yes X	3.8 → 1.3 X	1.3 X	Yes 4X
MAX, LT Fixed IU	Yes	Barely	Mid	Yes X	Barely	4 → 1.5 X	3.8 → 1.5 X	Yes 3X
MAX, LT Proportional IU	Yes	Yes	Low	Small X	Small X	>4 X	~3	Yes 3X
Σ, LT No IU	Yes	Yes	High	No	Small	1.7 X	1.3 X	Yes 2X
Σ, LT Fixed IU	Yes	No X	High	No	No	3.8	3.8	Yes 1X
MAX, NT No IU	Yes	Yes	Low	Yes X	Yes X	8 → 2.6 X	2.6	Yes 3X
MAX, NT Fixed IU	Yes	No X	Low	Yes X	Barely	8 → 3.0 X	8 → 3 X	Yes 4X
MAX, NT Proportional IU	Barely	Barely	Very low	Yes X	Yes X	10 → 7 X	~6 or 7 X	Yes 4X
Σ, NT No IU	Yes	Yes	Mid	No	Small	3.4	2.6	No
Σ, NT Fixed IU	Yes	No X	Mid	No	No	7 X	7 X	Yes 3X

Table 5. Summary of experimental results (bold) and model behaviors. The effects of spatial filtering and retinal inhomogeneity are excluded here, but considered later. As these factors affect the details of summation slopes, the third column here does not contribute to model rejection. All other qualitative mismatches between model and data are indicated by an X and lead to rejection. The number of Xs is tallied in the last column. IU: intrinsic uncertainty; LT: linear transducer; NT: nonlinear transducer.

was significant for all three observers in the experiment. The design effect can be reintroduced by allowing the number of irrelevant mechanisms to vary with stimulus size (i.e., involving proportional uncertainty, Figure 2e). This also has the benefit of increasing the slope of the psychometric function in the blocked design (Figure 4e) to something close to those in the human data (Figure 6b), but a design effect remains for the psychometric slope (Figure 4e) and is inconsistent with the results. Although rejected, this model is arguably the most successful of the MAX models (on a qualitative basis) and we revisit it again in Part IV along with the more conventional fixed uncertainty MAX model.

The linear transducer with linear summation and intrinsic uncertainty (Figures 3c and 5c) is rejected by its failure to predict the design effect for contrast sensitivity (Figure 6a), but might otherwise be considered a near miss. We also revisit this model in Part IV.

Another way to change the behavior of the MAX models is to weight the contribution of signal and noise according to the expected retinal inhomogeneity (Figure B1 in Appendix B). However, as we show in Appendix C (also look forward to Figure 7i), this predicts little or no improvement in contrast sensitivity beyond a stimulus diameter of four cycles, making that idea unpromising.

A successful toy model

The only model not rejected by a qualitative assessment of its behavior against our data was the noisy energy model (a nonlinear transducer followed by linear summation of signal and noise). The predicted levels of summation are broadly consistent with our experimental results: summation slopes are gentle, and there is a small effect of design in the correct direction (Figure 3b). Furthermore, the model predicts that there is no effect of area on the slope of the psychometric function for either condition and that $2.6 < \beta < 3.4$ (Figure 5c). This is all broadly consistent with the experimental results, though the empirical psychometric slopes are arguably a little high (average $\beta = 3.6$). This might be due to low levels of intrinsic uncertainty in the experiment that were not a part of this model. Instead, or as well as, it might be due to the small overestimation
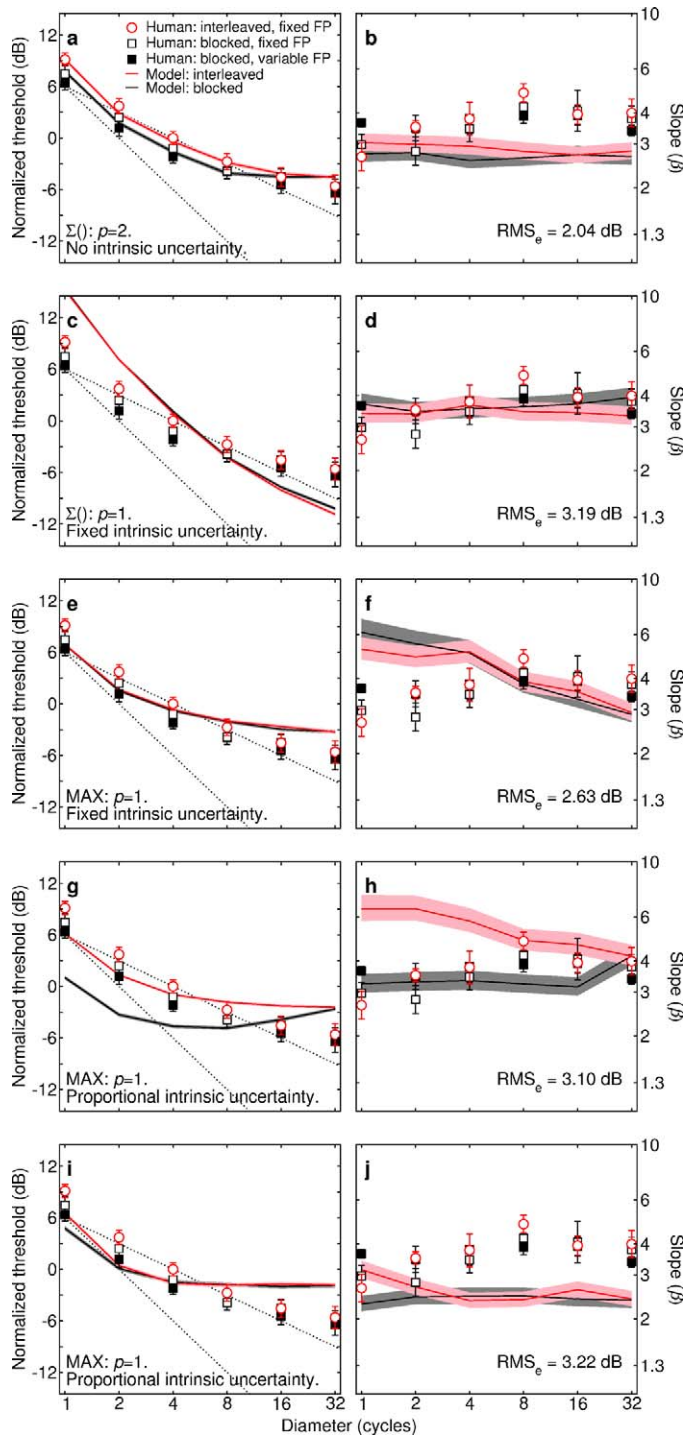
Figure 7. Predictions (a and b) and fits (c through j) of five models (different rows) to the average thresholds and psychometric slopes (different columns) replotted from Figure 6. The mean thresholds for models and data in the interleaved condition were normalized to 0 dB. The noisy energy model in (a and b) had no free parameters but produced the best predictions. The other four models each had a single free parameter, which was the level of uncertainty. RMS error (RMSe in decibels) was calculated in the conventional way (see Meese et al., 2007). For the slopes, this involved taking the log of the slope and multiplying by 20. This was somewhat arbitrary, but not critical for our conclusions. The

of the slope of the psychometric functions that is an inherent consequence of undersampling in typical psychophysical methods such as those used here (Wichmann & Hill, 2001; Wallis et al., in press).

Another minor failing is that the model predicts that the slope of the psychometric function should be slightly steeper for the interleaved design than the blocked design. However, the predicted difference is small compared to the variability in the experimental estimates of slope (Figure 6b), and is unlikely to be revealed by psychophysical experiments. We consider this model further in Part IV.

Results, Part IV: Quantitative model predictions and fits

Adding a front end to the models

The model predictions in Figures 2 through 4 do not include the effects of spatial filtering and retinal inhomogeneity. These are well-established properties of the visual system and have a marked effect on area summation (though little effect on the slope of the psychometric function). We introduce these effects here and implement them in several of the model variants from the previous section.

The images used in the experiment were sampled with a resolution of 12 pixels per grating cycle (though this was not critical) and multiplied by the so-called "witch's hat" attenuation surface shown in Figure B1 of Appendix B to simulate the effects of retinal inhomogeneity. This surface was derived from the mean parameters in Baldwin et al. (manuscript submitted for publication) and comprises different sensitivity losses for

←

RMSe combined across the two columns provided a single figure of merit for each of the four models (values in right hand column). In (c through j) we started with our best estimate of a suitable level of uncertainty and adjusted it in each direction in factors of two to find minima of the RMSe of the fits. For (c and d), $n = 1,200$. For (e and f), $n = 720,000$. For (g and h), $n = 36m$ for the interleaved condition (where $m = 117,032$) and $n = 36s$ for the blocked condition. For (i and j) $n = 9m$ and $9s$ for the interleaved and blocked conditions, respectively. In all of the simulations, $s = 156$ for the smallest stimulus and increased roughly in proportion to the square of the number of cycles. Deviations from this derive from the fact that each circular stimulus had an integer number of cycles but that added to this was a narrow boundary of lower contrast pixels (see Methods). The model calculations were performed across the region of the stimulus for which the envelope was greater than or equal to its half-height. This detail was not critical (see Appendix B).

the vertical and horizontal meridians in both the center and periphery. The attenuated image was then filtered by a cosine-phase Cartesian separable log-Gabor filter (Meese, 2010) with spatial frequency bandwidth of 1.6 octaves and orientation bandwidth of $\pm25°$ (bandwidths at half-height). The filter was matched to the spatial frequency and orientation of the target grating and its output was full-wave rectified, scaled to the range 0 to 1 (where unity was the maximum possible response) and multiplied by stimulus contrast (c). This front end of the model was used with several of the model variants described above.

Pooling was performed over the full width at half-height of the stimulus. This was done for simplicity and because we had no particular hypothesis about the strategy that would have been used by our observers (though see the next section). However, as we show in Appendix B, our model predictions were not critically dependent on this somewhat arbitrary decision.

Model predictions were normalized to the average thresholds in the interleaved designs and are shown in Figure 7 along with the experimental results replotted from Figure 6. By eye it is clear that the noisy energy model (Figure 7a and b) is the only one that provides an adequate account of the results. The predictions (curves; no free parameters) for the summation slopes (Figure 7a) are good, and those for the slopes of the psychometric functions (Figure 7b) are fair. The slight mismatches for the slopes are presumably due to the factors discussed in the previous section.

For completeness, we also considered the "best" runners-up identified in the qualitative analysis above. To do this, model fitting was performed where intrinsic uncertainty was a free parameter (see caption of Figure 7 for details of the fitting). When the transducer was fixed and summation was linear it was possible to get good fits against the slopes of the psychometric functions (Figure 7d) by setting a high level of intrinsic uncertainty. However, this model failed utterly with the summation slopes (Figure 7c), which are essentially independent of uncertainty.

High levels of intrinsic uncertainty are usually associated with the MAX summation rule (e.g., Pelli, 1985), and so we attempted fitting with either fixed or proportional uncertainty as the free parameter. Our MAX models were able to achieve bowed summation functions (Figure 6e and g), owing largely to the attenuation surface. However, there was too little or too much separation between the blocked and interleaved designs for the fixed and proportional uncertainty models, respectively. Furthermore, each of these MAX variants failed badly to describe the pattern of psychometric slopes (Figure 7f and h).

Note that the models in the second, third, and fourth rows of Figure 7 fail for exactly the reasons outlined in the qualitative analysis of Part III.

In Figure 7g, model performance in the blocked condition actually declined with an increase in stimulus area (i.e., there is an upturn to the black curve). This is because sensitivity in the periphery was so weak that the influence of the extra noise was the greater factor. Whether a suboptimal human MAXing observer should be expected to also perform in *this* suboptimal way is not clear. However, to provide the benefit of the doubt we reran the model with the noisy mechanisms weighted by a template constructed from the expected stimulus following spatial filtering and the witch's hat attenuation surface from Figure B1 in Appendix B. This weighting was also applied to the multiple "layers" of irrelevant noisy mechanisms. This down-weighted the more peripheral mechanisms, relevant and irrelevant alike. For the interleaved design, the template was that for the largest stimulus. The results are shown in Figure 7i and j. This strategy remedied the "upturn" problem described above but did nothing to improve the overall fit of the model. As we commented in the section *Near(ish) misses*, this weighting strategy (in conjunction with the MAX operator) is in stark contradiction to the human results since it predicts little or no benefit from stimulus area beyond a diameter of four stimulus cycles.

Finally, as noted by Tyler and Chen (2000, p. 3133) the absolute number of first-stage mechanisms involved can be important for the MAX operator. For example, in Figure 2a, the black curve for the blocked condition is not linear on double log scales, but concave, indicating that the size of benefit from doubling (say) the number of mechanisms that detect the signal will depend on the number of mechanisms involved in the first place (the benefit is greater for small numbers, over the initial range at least). In other words, knowledge of U alone is insufficient for the modeler; the number of mechanisms excited by the stimulus (s) must also be estimated. In Appendix C we develop the MAX model through several iterations of detail, and report on the effects of the absolute number of mechanisms. We are led to conclude that this consideration is unlikely to improve the fortune of the MAX model.

Overall then, the detailed quantitative modeling of this section (and Appendix C) confirms the preliminary qualitative analysis of the previous section: the noisy energy model performs well and there is little sign that the MAX model might be salvaged.

How to get template-matching to work

A class of model that is of interest to many researchers is the template-matching model. In this model, the observer multiplies the stimulus with (some variant of) a template of the expected stimulus (e.g., Burgess & Ghandeharian, 1984). These models have a
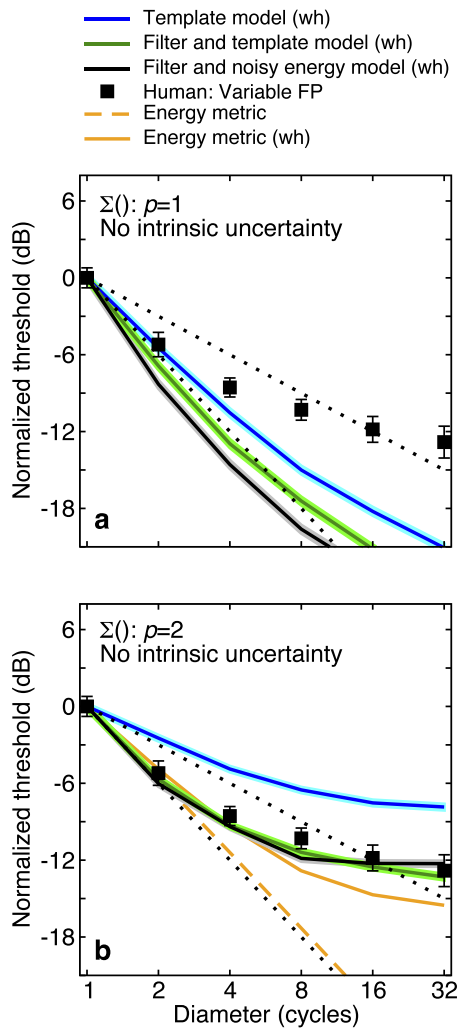
Figure 8. Development of a template-matching model. Data are for the blocked condition and replotted from Figure 6. Models and data are normalized to the smallest (left most) stimulus. (a) When the transducer was linear the summation functions were too steep for each of the three model variants. (b) When the transducer was a square-law, the template model (blue) was too shallow when the spatial filtering was omitted. With the spatial filtering in place, the template-matching model (green; RMS error = 0.67 dB) behaved in a very similar way to the noisy energy model (black; RMS error = 0.84 dB). Note that the black curve in (b) is replotted from that in Figure 7a. The dotted lines have slopes of −1/4 and −1/2 for comparison. The yellow curves are for energy metrics described in the main text.

long history and have received recent attention through the use of classification images (e.g., Tjan & Nandy, 2006; Neri, 2010; Murray, 2011). Our own noisy energy model is a form of simple template model in which the templates are matched to the area of the stimulus but are lacking detail about the luminance modulation across space. Of course, this is largely irrelevant to the study here because the area summation functions

depend on the way signal information is combined across multiple image cycles, not the processing within each cycle. Nevertheless, it is instructive to see how the ideal template-matching model (a cross-correlator) must be modified to fit our basic area summation results. The analysis in this section provides further support for our general conclusions and demonstrates the shortcomings of the cross-correlator. For simplicity, we consider only the results from the blocked experimental design, but the general approach could be extended to the interleaved design in exactly the same way as for our noisy energy model.

For clarity of exposition, models and data were normalized to the smallest stimulus. To keep the exposition simple, we ran the models using Monte Carlo simulations (as before), but analytic forms of the template models are easily derived. They produce predictions that are indistinguishable from those shown here (see Appendix D for the analytic version of our best version of the template model [green curve in Figure 8b]).

Cross-correlating observers know the signal exactly. We also assume that they know the details of the retinal attenuation surface (Appendix B). From this they construct a perfect template of the luminance modulation of the signal following retinal attenuation. Thus, the template is the product of the signal and the attenuation surface. This is used to compute a weighted sum of the signal and internal noise in each 2IFC interval, and the cross-correlating observer chooses the interval with the greatest response. The performance of this model is shown by the blue curve in Figure 8a. The benefit of stimulus size is lost to the real observers much more rapidly than it is to the cross-correlator. When our spatial filters were added to the model (green curve in Figure 8a) the mismatch between model and data became even worse. Similarly, when the transducer in the noisy energy model was linear ($p = 1$) this model also failed badly (black curve in Figure 8a). To try to remedy this problem we set the transducer to a square-law ($p = 2$) and also built the square law transduction into the template. This caused the model to underestimate the levels of summation in the human data (blue curve, Figure 8b). However, when our spatial filters were returned to the model and their effects were built into the expected template,[5] the predicted levels of summation became very much like the noisy energy model (green and black curves in Figure 8b). The increase in summation slope arises because the spatial filtering blurs the stimulus around its boundary, thereby reducing its energy. Because this effect is most severe for the smallest stimulus (where the boundary to area ratio is highest) this increases the initial part of the summation slope. A related factor is that the footprint of the filter-element (its receptive field) is larger than

the smallest stimulus. This means that the observer benefits from linear summation within that filter-element with the initial increase in stimulus area. This explanation applies to the template-matching model, the noisy energy model, and the MAX models in Figure 7 (see Appendix C and Meese [2010] for further comment).

Note that the template-matching model (green curve) performed very slightly better than the noisy energy model (black curve) (see figure caption for details). Presumably this is because the ideal template benefits from using very low weights for the insensitive parafoveal stimulus region, thereby attenuating the damaging effects of internal noise where the signal is weak. When the models are run on yet larger stimuli (not shown), the noisy energy model (black curve) curves back *upwards*, as negligible signal is added at the cost of recruiting further noise. This suboptimal strategy does not happen with the template-matching model (green curve), which asymptotes with larger stimuli (not shown).

Although we do not wish to claim that the visual system employs a detailed detection strategy as sophisticated as the template-matching model developed here, the use of a matched template (i.e., adjustable summation weights) provides the modeler with a convenient and parameter-free means by which to determine the summation region, avoiding our somewhat arbitrary decision to sum over the stimulus defined by the half-height of its envelope (though see Figure B2 in Appendix B for a defense of this).

The successful versions of the noisy energy model and the template-matching model in Figure 8b are similar models in many respects and owe their success to the properties that they share. In fact, if the template in the template-matching model were based on the envelope instead of the luminance profile, and if the noisy energy model were modified slightly to use that form of weighted summation, then the two models become identical (and still predict the experimental results [not shown]).

The dashed yellow curve in Figure 8b is the prediction for a contrast energy metric. It shows the expected improvement in sensitivity on the assumption that stimulus energy is constant at detection threshold (see appendix E of Meese [2010] for implementation details). It is very similar to the fiducial contour of $-1/2$ (lower black dotted line), differing slightly due to the narrow (2 pixel) blurred skirt added to the surround of our stimuli. The solid yellow curve is the same energy metric but with the attenuating effects of retinal sensitivity taken into account (i.e., the stimulus was multiplied by the attenuation surface in Figure B1 in Appendix B). This prediction is clearly bettered by each of the filter-models (black and green curves), though it

does predict the variation in sensitivity fairly well over the first three stimuli.

Discussion

Main findings

We measured thresholds and the slopes of the psychometric functions using interleaved and blocked experimental designs for centrally placed circular patches of grating with six different diameters. We confirmed the well-known finding that the improvement in sensitivity with area decelerates with area. But we found sensitivity was slightly higher for the blocked experimental design, particularly for the small stimuli. Furthermore, we found no evidence for a systematic effect of experimental design on the slopes of the psychometric function nor any evidence for a decrease in the slope of the psychometric function with stimulus area, which were fairly constant in the region of $\beta = 3$ to 4. We developed 10 canonical models of the summation process involving linear and nonlinear transducers, various forms of uncertainty, and linear sum and MAX pooling operations. Of these, only the noisy energy model made the correct qualitative predictions. When this was extended to include spatial filtering and retinal inhomogeneity, it produced fairly good quantitative predictions of our results. We found no variant of the MAX model (a contemporary implementation of probability summation) that was able to account for all of the results. A cross-correlator (matched template) model also produced good predictions, but only when it included the same key features as the final version of the noisy energy model: retinal inhomogeneity, spatial filtering, square-law contrast transduction, and integration of signal and noise over stimulus area. In short, the template is not a substitute for spatial filtering but comes after it.

Some of our conclusions share similarities with those in a recent study by Neri (2010). In that work, the task was detect a single vertical bright bar target placed in a background of distractor bars of random intensity. Neri (2010) did not manipulate target area but did manipulate uncertainty by cuing the potential target region to various spatial extents. He concluded that detection did not involve a MAX operation over space but linear spatial summation following approximately square-law transduction.

Summation above threshold

The long-range summation process that we propose would be suitable for representing image structures that

extend beyond the footprint of a single receptive field. However, the responses of these integrators would need to be kept in check above threshold by a suitable hierarchy of contrast gain control if they are to also carry a code for luminance contrast (Meese & Summers, 2007). These ideas were explored, tested, and confirmed by Meese and Baker (2011).

Why the probability summation model fails

Our aim was to investigate whether area summation of contrast derives from linear spatial integration (possibly following nonlinear contrast transduction) or spatial probability summation (implemented by a MAX operator following noise). Our approach was to bring an additional constraint to the traditional area summation experiment by manipulating extrinsic uncertainty. The finding that sensitivity was higher for the blocked design than the interleaved design implies that we were successful in achieving this. In other words, whatever the unknown level of the intrinsic uncertainty, it was sufficiently low for the experimental manipulation of extrinsic uncertainty to contribute to total uncertainty. This meant that for the MAX operator, the level of total uncertainty always decreased with stimulus size for the interleaved design and that predicted that the slope of the psychometric function should become shallower with area. That prediction was not consistent with our experimental results. Some of these problems for the MAX operator were overcome by including the variation in retinal sensitivity in a weighting template (see Figure 7i and j). However, this wrongly predicted that sensitivity should improve only over the first three stimulus sizes.

If area summation of contrast is linear why has the probability summation model ruled for so long?

A striking conclusion from the study here is that summation of contrast over area is linear. This might be counterintuitive, since summation slopes (in model and data) do not look linear (i.e., they do not have a slope of -1)—even when the effects of retinal inhomogeneity are removed (e.g., see Figure 2). This is because the combined effects of square-law contrast transduction and noise integration produce a fourth-root summation rule (Meese, 2010). This rule is broadly consistent with the levels of summation found in previous studies (e.g., Robson & Graham, 1981) and also the back-pocket (i.e., casual) model of probability summation (e.g., Rohaly et al., 1997; Meese et al., 2005). However, the analysis here (Appendix A; see also Tyler & Chen, 2000) shows that the predictions for

probability summation are not as simple as the back-pocket model assumes—they depend on several factors including the details of uncertainty and the transducer exponent. Thus, our contention is that the tendency for empirical summation functions to approximate a fourth-root rule has led to them being misinterpreted in terms of probability summation, even though well-formulated models show that probability summation does not necessarily produce a fourth-root rule—see Figure 2. The headline here is that when a signal detection formulation of the probability summation model (involving the MAX rule) is put up against stringent tests involving the manipulation of uncertainty (blocked versus interleaved experimental design) and the measurement of the slope of the psychometric function, it fails, in this study, miserably.

The stimulus energy metric: Another deception?

It is sometimes claimed that basic energy metrics perform quite well, particularly over smaller stimulus ranges (e.g., Watson & Ahumada, 2005), and the analyses here (yellow curves in Figure 8b) appear to support that. However, using contrast modulated "Battenberg" patterns, Meese (2010) showed that in fact, the stimulus energy metric is a very poor predictor of contrast sensitivity. For example, when local orientation was modulated, keeping the stimulus energy fixed, sensitivity varied over a range of about 4 dB. The energy metric failed badly in that study because it makes no allowance for spatial summation *within* filter-elements over short distances. So why has it done so well for the smaller stimuli here? When the filters are removed from our successful models, the initial slope is shallow—approximately a fourth-root slope owing to the cascade of square-law transduction and integration of noise (e.g., the blue curve in Figure 8b). However, when the filters are returned, the slope is steepened owing to the effects of linear summation within the filter-elements (compare the blue and green curves in Figure 8b) and the result happens to be close to a quadratic summation rule (slope $=-1/2$). Thus, our contention is that just as the combined effects of transduction and noise can masquerade as probability summation over larger distances, when these processes are combined with short-range linear summation *within* filters, they can masquerade as a stimulus energy metric over shorter distances. The filter-based models that we propose predict both of these deceptions.

Uncertainty and the nonlinear transducer

Previous studies have found that spatial uncertainty affects the slope of the psychometric function in

contrast detection (Shani & Sagi, 2005) and contrast discrimination experiments for multiple target patches (Meese, Hess, & Williams, 2001; Nachmias, 2002) and that phase uncertainty affects the dipper region for stereo discrimination thresholds (Georgeson, Yates, & Schofield, 2008), all broadly consistent with theory (Pelli, 1985). However, this is the first study to show that extrinsic uncertainty (controlled by experimental design) can affect detection thresholds for stimulus patches of various sizes. Uncertainty is clearly an important part of modeling contrast detection, and there has been a long-standing debate about whether early visual nonlinearities (e.g., Nachmias and Sansbury, 1974) should be attributed to substantial levels of uncertainty (Pelli, 1985) or a nonlinear transducer (Legge & Foley, 1980). We have not been able to find a model variant involving a linear transducer and high levels of uncertainty that can account for the results here, so our study contributes to this debate, coming down in favor of a nonlinear (square-law) transducer. This is consistent with the requirements of a contrast energy computation (e.g., Watson, Barlow, & Robson, 1983; Manahilov et al., 2001; Meese, 2010). Meese and Summers (2009) arrived at a similar position, concluding that the levels of intrinsic uncertainty are modest for the contrast detection of patches of sine-wave grating (e.g., $U_{int} \approx 3$).

Limitations

Although we have considered a wider range of models of area summation than has ever been done previously, our approach is not without its limitations. It might be criticized for constraining our various models by our choices of fixed parameters. It was necessary to do this to reduce the complexity of the modeling process to a manageable level. Furthermore, we tried to salvage the failing models by adjusting some of the parameter values (see previous sections and Appendix C). This was to no avail, but we do acknowledge that it might be possible to model our results with a model refinement that has not been considered here. For example, one possibility is that the MAX operator uses a weighting template that has a different form from the retinal sensitivity surface. Another possibility is that linear summation operates over a limited range, with the strategy switching outside that range. In fact, we acknowledge that the work here does not lead us to claim that summation must be linear over the entire stimulus region. Other recent work suggests that the upper limit might be somewhat less than our maximum stimulus size here (Baker & Meese, 2011). It is also possible that uncertainty involves some combination of fixed and proportional factors. No doubt, there are also other possibilities that we have not thought of.

Another limitation of our work is that we have restricted our analysis to the case of additive noise. It is possible that the visual system contains a component of multiplicative noise (Tolhurst, Movshon, & Thompson, 1981) and this has implications for models of probability summation (Tyler & Chen, 2000). However, the link between the neurophysiology and the threshold psychophysics is not well established in this regard, and what relevant psychophysical work there is does not help to decide (Georgeson & Meese, 2006). Therefore, we chose to leave this matter at rest until a firmer position can be established, particularly since the neurophysiological evidence (Gur, Beylin, & Snodderly, 1997) is not as clear cut as often supposed (see Georgeson & Meese [2006] for discussion).

Finally, our exposition of the noisy energy model (and the matched-template model) involves a discrete set of pooling mechanisms (i.e., templates). However, we cannot conclude at this stage whether these mechanisms are hard-wired, or whether the observer constructs them as needed to match the various targets.

Conclusions

There is a long-standing belief in the psychophysical literature that spatial integration of image contrast stops at the level of spatial filtering, typical of that found in simple cells in primary visual cortex. On this view, physiological summation of contrast (*signal combination*) extends over a couple of stimulus cycles, at best. The psychophysical work here, and elsewhere (Kersten, 1984; Mayer & Tyler, 1986; Manahilov et al., 2001; Foley et al., 2007; Meese & Hess, 2007; Meese & Summers, 2009; Meese, 2010; To, Baddeley, Troscianko, & Tolhurst, 2010; Meese & Baker, 2011) suggests that signal combination is spatially more extensive than this.

Our work also favors the nonlinear transducer model (Legge & Foley, 1980) over the uncertainty model (Pelli, 1985), though low levels of intrinsic uncertainty are not ruled out (e.g., Meese & Summers, 2009).

Our results are also well predicted (no free parameters) by a matched template model, but only when the template follows first-order spatial filtering and square-law contrast transduction and the template is subject to the spatial integration of internal noise in addition to the signal.

Finally, our theoretical analysis (Appendix A) shows that, in general, it is not appropriate to use Minkowski summation as an approximation to probability sum-

mation where the Minkowski exponent is set according to the slope of the psychometric function.

Acknowledgments

Footnotes

[1] The second dogma of spatial vision pertains to suprathreshold summation (Meese & Summers, 2007; Meese & Baker, 2011) and is not relevant here.

[2] In this paper we use the term 'retinal inhomogeneity' quite loosely. We mean it to imply the variation of contrast sensitivity across the retina regardless of whether the origins of that variation are sub-cortical or cortical.

[3] From Pelli (1985), Weibull β increases very nearly linearly over five orders of magnitude of log U. The threshold function is slightly more compressive over a similar range. However, the log of threshold is markedly compressive when plotted against the log of U. In other words, once uncertainty is very high, enormous amounts of extra uncertainty are required for it to influence log sensitivity appreciably.

[4] The ideal strategy is to search for the maximum difference of each of the six normalized pooling mechanisms across the 2IFC interval. However, it seems unlikely that the observer would retain all of the necessary information from the first interval, and so we opted for applying the MAX operator across the six normalized responses within each interval. For the conditions considered here these two strategies produced negligible differences (not shown).

[5] We built the effects of filtering and nonlinear transduction into the expected template because for an observer with this front-end, that is the ideal strategy. However, further simulations in which the template was that of the stimulus without filtering and transduction showed that this detail was not important for achieving the good model prediction shown here.

References

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. [PubMed] [Article]

Baldwin, A. S., Meese, T. S., & Baker, D. H. The attenuation surface for contrast sensitivity is a bilinear 'witch's hat' within the central visual field. *Journal of Vision*, in press.

Bird, C. M., Henning, G. B., & Wichmann, F. A. (2002). Contrast discrimination with sinusoidal gratings of different spatial frequency. *Journal of the Optical Society of America A, 19*(7), 1267–1273.

Bonneh, Y., & Sagi, D. (1999). Contrast integration across space. *Vision Research, 39*, 2597–2602.

Burgess, A. E., & Ghandeharian, H. (1984). Visual signal detection. II. Signal-location identification. *Journal of the Optical Society of America A, 1*(8), 906–910.

Cadieu, C., Kouh, M., Pasupathy, A., Conner, C. E., Riesenhuber, M., & Poggio, T. (2009). A model of V4 shape selectivity and invariance. *Journal of Neurophysiology, 98*, 1733–1750.

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

Finn, I. M., & Ferster, D. (2007). Computational diversity in complex cells of cat primary visual cortex. *Journal of Neuroscience, 27*, 9638–9647.

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.

García-Pérez, M. A. (1988). Space-variant visual processing: Spatially limited visual channels. *Spatial Vision, 3*, 129–142.

García-Pérez, M. A., & Alcalá-Quintana, R. (2007). The transducer model for contrast detection and discrimination: formal relations, implications, and an empirical test. *Spatial Vision, 20*, 5–43.

Georgeson, M. A., & Meese, T. S. (2006). Fixed or variable noise in contrast discrimination? The jury's still out... *Vision Research, 46*, 4294–4303.

Georgeson, M. A., Yates, T. A., & Schofield, J. (2008). Discriminating depth in corrugated stereo surfaces: Facilitation by a pedestal is explained by removal of uncertainty. *Vision Research, 48*, 2321–2328.

Graham, N., Robson, J. G. & Nachmias, J. (1978).

Grating summation in fovea and periphery. *Vision Research, 18*, 815–825.

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

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

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

Gur, M., Beylin, A., & Snodderly, D. M. (1997). Response variability of neurons in primary visual cortex (V1) of alert monkeys. *Journal of Neuroscience, 17*, 2914–2920.

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

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

Laming, D. (1988). Precis of sensory analysis. *Behavioural and Brain Sciences, 11*, 275–339.

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

Manahilov, V., & Simpson, W. (1999). Energy model for contrast detection: Spatiotemporal characteristics of threshold vision. *Biological Cybernetics, 81*, 61–71.

Manahilov, V., & Simpson, W. (2001). Energy detection for contrast detection: Spatial-frequency and orientation selectivity in grating summation. *Vision Research, 41*, 1447–1560.

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

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

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., & Baker, D. H. (2011). Contrast

integration of 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. [PubMed] [Article]

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

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

Meese, T. S., Hess, R. F., & Williams, C. B. W. (2001). Spatial coherence does not affect contrast discrimination for multiple Gabor stimuli. *Perception, 30*, 1411–1422.

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

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

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.

Meinhardt, G. (2000). Detection of compound spatial patterns: Further evidence for different channel interactions. *Biological Cybernetics, 82*, 269–282.

Mortensen, U. (1988). Visual contrast detection by a single channel versus probability summation among channels. *Biological Cybernetics, 59*(2), 137–147. doi:10.1007/BF00317776.

Mortensen, U. (2002). Additive noise, Weibull functions and the approximation of psychometric functions. *Vision Research, 42*, 2371–2393.

Murray, R. F. (2011). Classification images: A review. *Journal of Vision, 11*(5):2, 1–25, http://www.journalofvision.org/content/11/5/2, doi:10.1167/11.5.2. [PubMed] [Article]

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

Nachmias, J. (2002). Contrast discrimination with and

without spatial uncertainty. *Vision Research, 42*, 41–48.

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: Discrimination may be better than detection. *Vision Research, 14*, 1039–1041.

Neri, P. (2010). Visual detection under uncertainty operates via an early static, not late dynamic, nonlinearity. *Frontiers in Computational Neuroscience, 4*(151), 1–17.

Párraga, C. A., Troscianko, T., & Tolhurst, D. J. (2005). The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and multi-resolution model. *Vision Research, 45*, 3145–3168.

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

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

Rashbass, C. (1970). The visibility of transient changes of luminance. *Journal of Physiology, 210*, 165–186.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology, 12*, 162–168.

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

Rohaly, A. M., Ahumada, A. J., & Watson, A. B. (1997). Object detection in natural backgrounds predicted by discrimination performance and models. *Vision Research, 37*, 3225–3235.

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

Rovamo, J., Mustonen, J., & Nasanen, R. (1994). Modeling contrast sensitivity as a function of retinal illuminance and grating area. *Vision Research, 34*, 1301–1314.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America, 61*, 1176–1186.

Serre, T., Woilf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 1–17.

Shani, R., & Sagi, D. (2005). Eccentricity effects on lateral interactions. *Vision Research, 45*, 2009–2024.

Summers, R. J., & Meese, T. S. (2009). The influence of fixation points on contrast detection and discrimination of patches of grating: Masking and facilitation. *Vision Research, 49*, 1894–1900.

Tjan, B. S., & Nandy, A. S. (2006). Classification images with uncertainty. *Journal of Vision, 6*(4):8, 387–413, http://www.journalofvision.org/content/6/4/8, doi:10.1167/6.4.8. [PubMed] [Article]

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research, 41*, 414–419.

Tootle, J. S., & Berkley, M. A. (1983). Contrast sensitivity for vertically and obliquely oriented gratings as a function of grating area. *Vision Research, 23*, 907–910.

To, M. P. S., Baddeley, R. J., Troscianko, T., & Tolhurst, D. J. (2010). A general rule for sensory cue summation: Evidence from photographic, musical, phonetic and cross-modal stimuli. *Proceedings of the Royal Society B, 278*, 1365–1372.

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.

Wallis, S. A., Baker, D. H., Meese, T. S., & Georgeson, M. A. The psychophysical function in spatiotemporal contrast vision: A method for decoupling non-stationarity from the psychometric slope. *Vision Research*, in press.

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

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

Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What does the eye see best? *Nature, 302*, 419–422.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics, 63*, 1293–1313.

Wilson, H. R. (1980). A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics, 38*, 171–178.

Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research, 19*, 19–32.

Appendix A. Probability summation and the slope of the psychometric function

Here we show the relation between the slope of the psychometric function (the Weibull parameter β) and the Minkowski exponent, γ, needed to produce the level of summation predicted by each of our model variants. The analysis here involves assessing the level of summation that is predicted when the number of equally sensitive mechanisms is doubled at various points along each of the model functions, where Minkowski summation is given by:

$$resp_{overall} = \left[\sum_i (|r_i|^\gamma) \right]^{1/\gamma}. \qquad (A1)$$
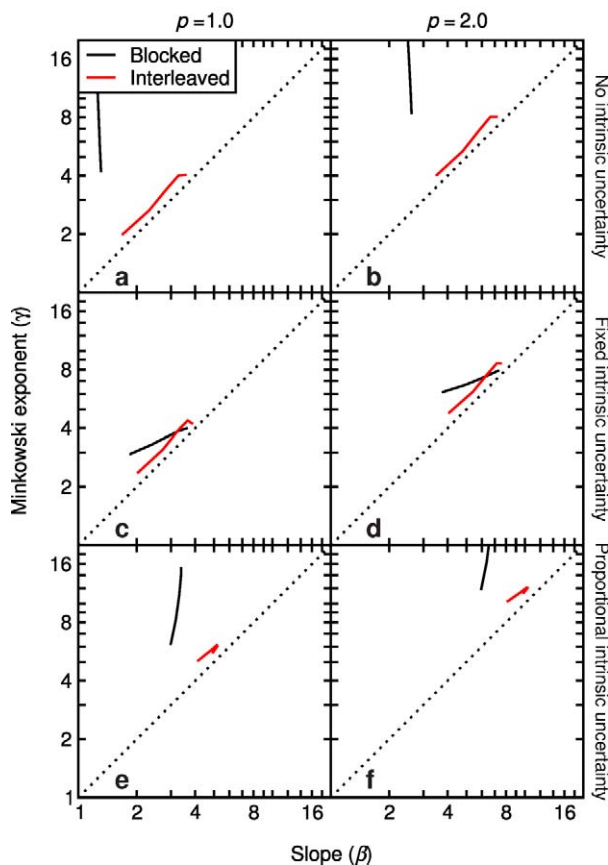
Figure A1. Relation between the slope of the psychometric function (β) and the Minkowski exponent (γ) needed to produce the levels of summation predicted by several variants of the probability summation model when implemented by MAX pooling. Different columns are for different transducer exponents ($p = 1$ or 2) and different rows are for the different forms of intrinsic uncertainty used in our canonical models (see right hand labels).

To approximate our simulations with continuous functions, we began by fitting each summation curve in Figure 2 with a cubic equation, $thresh(x) = Ax^3 + Bx^2 + Cx + D$ (in decibels), and each of the curves describing the slope of the psychometric function in Figure 4 with a quadratic equation, $psychSlope(x) = Ex^2 + Fx + G$. In each case, $x = \log_2(i)$ where i is the number of equally excited mechanisms.

The Minkowski exponent (γ in Equation A1) was estimated from the summation curves for integer increments of x as follows:

$$\gamma(x) = \frac{1}{\left(\log_2 \left\{ 10^{\left[\frac{thresh(x) - thresh(x+1)}{20} \right]} \right\} \right)}.$$

For each $\gamma(x)$, the associated value of β was given by

$$\beta(x) = [psychSlope(x) + psychSlope(x+1)]/2.$$

For the MAX models and the levels of uncertainty considered here, Figure A1 shows how the Minkowski exponent should be set to achieve the appropriate levels of model summation for a given estimate of the slope of the psychometric function (β). In the classical analysis of probability summation, the slope of the psychometric function is treated as an estimate of the Minkowski exponent (Quick, 1974; Robson & Graham, 1981). In this view, $\gamma = \beta$, as indicated by the diagonal dotted lines in Figure A1. Our analysis shows that this equivalence is never actually met. If intended as an approximation to probability summation, then the Minkowski exponent (γ) should always be set higher than the slope of the
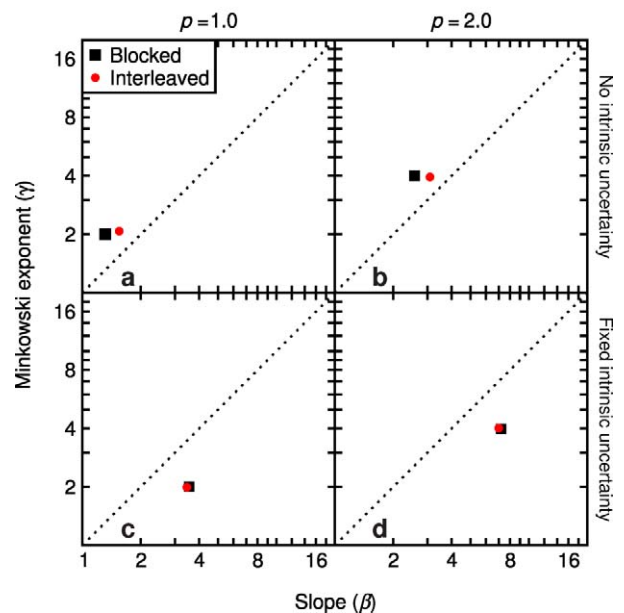

Figure A2. Similar to Figure A1 but for pooling by linear summation.

psychometric function (β). In some cases the difference is marginal, but in others it is substantial. But choosing an appropriate value is likely to be difficult. In some cases the functions are almost vertical (e.g., see the black curves in panels a, b, e, and f) meaning that very small changes in the estimate of the slope of the psychometric function lead to large changes in the Minkowski exponent. Furthermore, the functions are different for the blocked and interleaved designs because of the different effects of uncertainty. More troublesome still, the relationship between γ and β depends on the nature of the intrinsic uncertainty, a parameter over which the experimenter has little or no control.

For completeness, Figure A2 shows the results of the analysis applied to the four variants of linear pooling that we considered. Note that the Minkowski exponent is given directly by $2p$. For the blocked condition, the slope of the psychometric function is given by $\sim 1.3p$ in the absence of uncertainty. For the interleaved condition, it is a little higher. In all cases, uncertainty increases the slope of the psychometric function but leaves the Minkowski exponent untouched.

Appendix B. The front end of the detection models in Part IV

Images were sampled with a resolution of 12 pixels per carrier cycle (though this was not critical) and

multiplied by the attenuation surface shown in Figure B1, to simulate the effects of retinal inhomogeneity. This surface was derived from the mean parameters in Baldwin et al. (manuscript submitted for publication) and comprises different rates of sensitivity loss for the vertical and horizontal meridians. The decline in sensitivity with eccentricity is also greater over approximately the first eight cycles than it is subsequently. The attenuated image was then filtered by a cosine phase Cartesian separable log-Gabor filter (Meese, 2010) with spatial frequency bandwidth of 1.6 octaves and orientation bandwidth of $\pm 25°$ (bandwidths at half-height; see inset on Figure 2). (The use of a cosine phase filter was not critical, sine filters or the quadrature pair produced very similar results.) The filters were matched to the spatial frequency and orientation of the carrier grating and their outputs were full-wave rectified and scaled to the range 0 to 1 (where unity was the maximum possible response). Each pixel was then treated as r_i in the relevant equation from the section on the toy models in the main body of the report. The signal region was defined by the half-height of its envelope (i.e., it has a diameter of ($12 \times$ cycles $+ 2$) pixels, where *cycles* is the number of carrier cycles in the plateau region of the full stimulus). For the stimuli used here, this was a convenient if arbitrary solution to the problem of defining the summation region. However, Figure B2 shows that the model predictions are not critically dependent on this parameter. For example, other reasonable choices such as summing
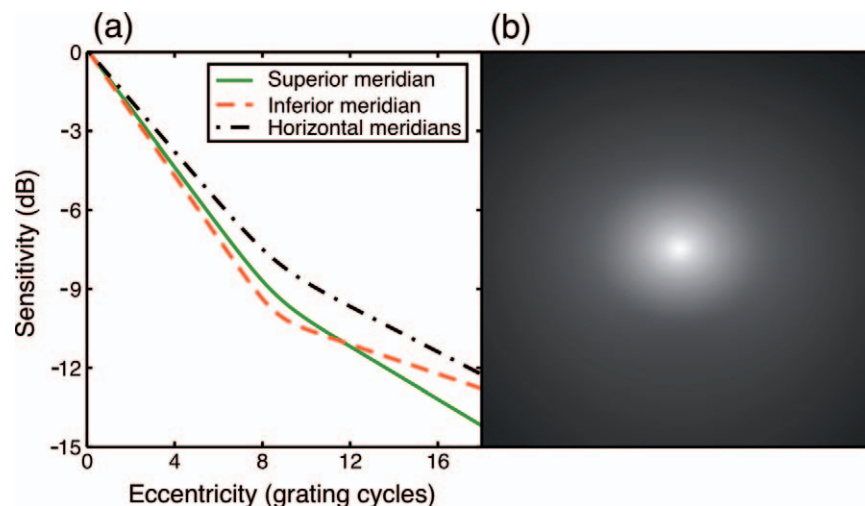

Figure B1. The "witch's hat" attenuation surface used in the modeling in Part IV and replotted from Baker and Meese (2011). The surface was derived by Baldwin et al. (manuscript submitted for publication) who measured sensitivity to 4 c/deg patches of grating over the central retina (a diameter of 9°). Each grating patch was surrounded by a low contrast circular ring to lessen potential effects of uncertainty. (a) Log sensitivity was a bilinear function of eccentricity. The functions for the left and right horizontal hemi-meridians were the same, but slightly different from each of the functions for the superior and inferior vertical meridians (see Baldwin et al. [manuscript submitted for publication] for equations and matlab code using the "average" parameter values in their table 4). For the largest stimuli used in the experiments here (a radius of 16 cycles), model sensitivity declined by about 12 dB (about a factor of 0.25) from the center to the edge of the stimulus. (b) A gray-level image of the attenuation surface.
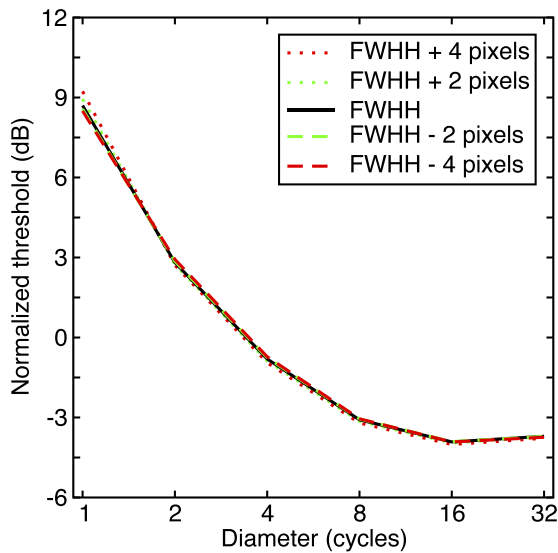
Figure B2. Effects of template diameter on the predictions for the noisy energy model for the blocked experimental design (curves are normalized to their means). The black curve is for summation across the full-width at half-height (FWHH) of the stimulus and was used in the main body of the report. The other curves are for where the diameter was extended or reduced by 2 or 4 pixels, as shown in the legend.

Figure B3. Effects of the exponent *p* (nonlinear transducer) on the predictions for the noisy energy model for the blocked experimental design (curves are normalized to their means).

over only the central plateau (green dashed curve) or the entire stimulus (green dotted curve) produce negligible changes to the predictions.

Additional pixels containing no signal were included to simulate the effects of uncertainty where appropriate.

We considered only a single filter (with multiple filter-elements, or receptive fields) in the modeling here (i.e., one tuned to a single orientation and spatial frequency). Where stimulus energy is lost to this filter (e.g., the smallest stimulus; see the section *How to get template-matching to work* in the main body of the report) one might expect that it would be picked up by other filters in the human brain, not included in the model. However, including the responses of such filters in the template will not fully compensate the loss because (a) extra noise will also be recruited from each extra filter-element and (b) when the transducer is nonlinear ($p > 1$), the impact of stimulus energy is diminished when it is spread across multiple filter-elements. For simplicity, we chose to not include additional filters here.

We did not try to optimize the fits by allowing the transducer exponent *p* to be a free parameter. However, the effects of varying this parameter are shown in Figure B3. Clearly, $p = 2$ (black curve) is close to optimal, though the upturn to the right of the functions, readily seen where $p > 2$, could be remedied if the signal and noise were weighted by the retinal attenuation surface (Figure B1) before summation (see
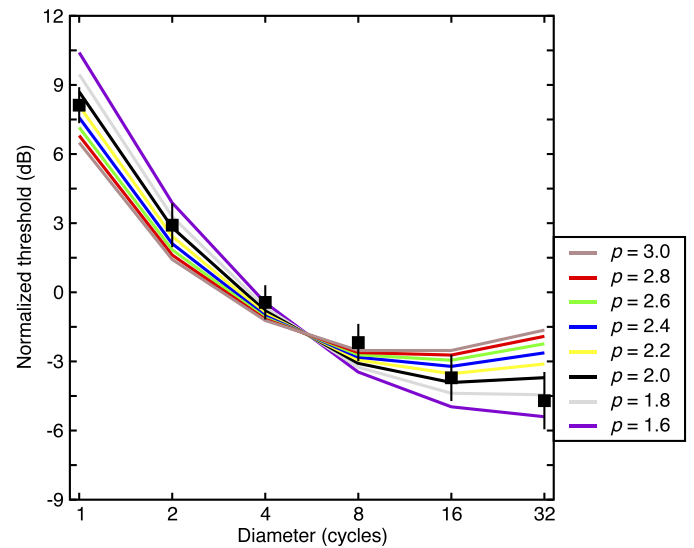
the section *How to get template-matching to work* in the main body of the report).

Appendix C. Variants of the MAX model

Here we develop the MAX model for interleaved and blocked experimental designs to illustrate the impact (or lack of it) that various model details have on its behavior. The predictions for thresholds and slopes of the psychometric functions are shown in Figure C1.

In the first row (Figure C1a and b) we begin with a variant that is very similar to the toy model from Figure 2a. The only difference is that in the original toy models the smallest stimulus excited just a single mechanism whereas here, the number of mechanisms was (somewhat arbitrarily) set by the number of pixels in the stimulus over the full width at half height of its envelope. For the smallest stimulus, this was 156 mechanisms. As anticipated from the results in Figure 2a (see also Tyler & Chen [2000] and our comments in the section *Adding a front end to the models* in the main body of the current report), this had the mere effect of slightly reducing the predicted level of summation from that seen in the earlier toy model (i.e., the black curve has a slightly shallower slope in Figure C1a than it does in Figure 2a). In other simulations (not shown) we set the number of mechanisms to four for the smallest stimulus, consistent with a sampling regimen of two samples per cycle as requested by a reviewer. This produced a summation slope intermediate to the other two but had little other
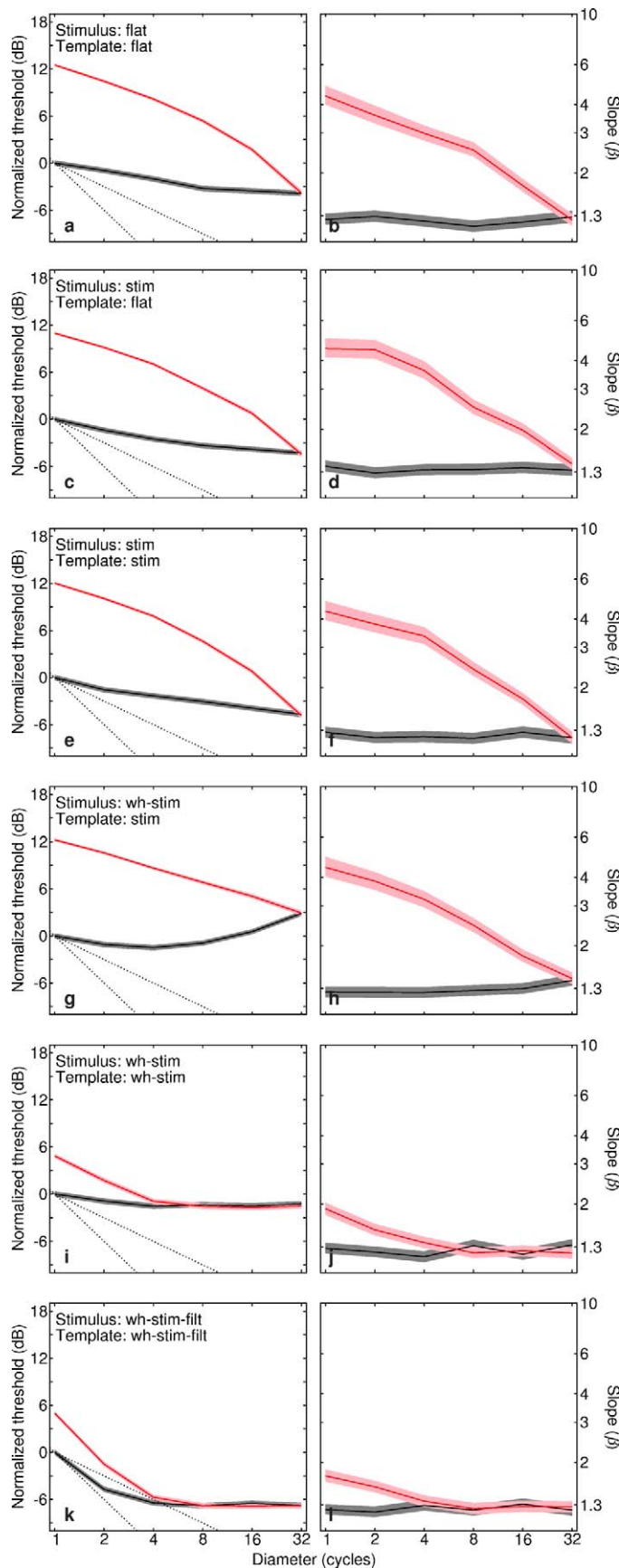
influence.

In the second row (Figure C1c and d) we added in the luminance modulation of the stimulus (used in the experiment) and full-wave rectified it. This had very little effect on the predictions. In the third row (Figure C1e and f) we weighted the mechanism responses with a template matched to the stimulus. This had very little effect on the predictions. In the fourth row, we added the witch's hat attenuation surface (Figure B1) to the front end of the model. This had very little effect on the general form of the predictions, but it did cause an upturn in the threshold predictions for the blocked design (black curve) towards the right. This effect was abolished when we added the effects of the witch's hat to the weighting template in the fifth row (Figure C1i and j). This also diminished the design effects considerably (the difference between the red and black curves) and essentially abolished the effects of stimulus size beyond a diameter of four stimulus cycles. Part of the reason for these changes is that the down-weighting of the numerous insensitive peripheral mechanisms effectively reduced the level of uncertainty in the interleaved design (red curves). In the sixth row (Figure C1k and l) we added spatial filtering (see Appendix B) to both the stimulus and the template. This made the initial part of the summation slopes steeper (owing to linear summation *within* the filter-elements; see also Meese [2010]) for both designs (Figure C1k) but had little or no effect on the slopes of the psychometric functions (Figure C1l).

Appendix D. Analytic expression for the template-matching model

Here we provide analytic expressions for two slightly different versions of the template-matching model for the blocked experimental design. Here and in previous work we have applied the witch's hat

←

Figure C1. Predictions for thresholds (left column) and slopes of the psychometric function (right column) through several stages of development of the MAX model. Red and black curves are for interleaved and blocked designs, respectively. The contrast transducer was linear. Panel headings: flat: the stimulus and/or template was uniform. stim: the stimulus and/or template was a full-wave rectified sine-wave grating modulated by the raised cosine window function used in the experiment (i.e., with a blurred region two pixels wide). wh: the stimulus and/or template were multiplied by the witch's hat attenuation surface in Figure B1. filt: the stimulus and template were subject to the spatial filtering described in Appendix B. See text for further details.

(Figure B1) to our stimuli before filtering, mainly for reasons of technical convenience. However, applying the witch's hat after filtering produces almost identical results (for the type of stimuli used here) and permits an analytic expression that can be discussed with greater insight. We present that version first (stimulus → filter → witch's hat → transduction), then the version equivalent to the stochastic model used in Figure 8b (stimulus → witch's hat → filter → transduction).

For convenience, we collapse the two-dimensional stimulus space into a single dimension with index i and length (number of first-stage mechanisms) s. We refer to the filtered stimulus as *stim*, the witch's hat as *witch*, and stimulus contrast as c. The signal to noise ratio (*SNR*) in the target interval is given by:

$$SNR = \frac{c^2 \sum_{i=1:s}(witch_i^4 stim_i^4)}{\sqrt{\sum_{i=1:s}(witch_i^4 stim_i^4)}}. \tag{D1}$$

The c term is squared because of the square-law nonlinear transduction. The $witch_i$ and $stim_i$ terms on the numerator are squared once by the nonlinear transduction, then again because the expected signal is multiplied by an exact template of itself (i.e., the template is also subject to filtering, transduction, and the witch's hat). These terms are then summed linearly, as for a cross-correlator. Note that for the stimuli used here, the *witch* term is responsible for the concave bowing of the summation function (e.g., Figure 8b), whereas the *stim* term has no effect on the form of the summation function other than it carries the effects of spatial filtering, discussed in the section *How to get template-matching to work* in the main body of the report.

The $witch_i$ and $stim_i$ terms are squared once on the denominator because the template is subject to square-law transduction and is used to weight the noise terms at each location i. These squared terms are standard deviations and must be squared again to give the local variances, which are summed, and the square root delivers the standard deviation of the overall noise term. Thus, the *SNR* is the ratio of the weighted linear sum of the signals squared (owing to nonlinear transduction) and the weighted linear sum of the noise sources.

Note that if the witch's hat were a flat uniform surface, and the modulation transfer function of the filter were flat, then the summation slope would be fourth root (−1/4) on log-log axes, determined entirely by the c^2 term and the square root influence of the accumulated noise. The

fourth-power terms in Equation D1 are irrelevant to its fourth-root summation behavior!

For the model in Figure 8b, the witch's hat was applied to the stimulus before filtering. In this case, the analytic expression becomes

$$SNR = \frac{c^2 \sum_{i=1:s}(template_i^4)}{\sqrt{\sum_{i=1:s}(template_i^4)}}, \tag{D2}$$

where *template* is the stimulus following attenuation by the witch's hat and spatial filtering.

Appendix E: Stimulus equation

The equation for the stimuli used in the experiments here was as follows:

$$stim_{xy} = L_0 + L_0 win_{xy}\{c.sin[2\pi(y - 256.5)/cycpix]\}, \tag{E1}$$

where x,y ($1 \leq x,y \leq 512$) are indices into a two-dimensional (512 × 512) pixel array, L_0 is mean luminance, c is the Michelson contrast of the carrier (and stimulus), *cycpix* is the number of pixels per cycle (=12), and win_{xy} is an envelope function, defined as follows:

$$IF\sqrt{\left[(x - 256.5)^2 + (y - 256.5)^2\right]}$$
$$\leq platpix\ THEN\ win_{xy} = 1 \tag{E2}$$

$$ELSEIF\sqrt{\left[(x - 256.5)^2 + (y - 256.5)^2\right]}$$
$$> (platpix + skirtpix)\ THEN\ win_{xy} = 0$$

$$ELSE\ win_{xy}$$
$$= \left[sin\left(\pi\left\{\sqrt{\left[(x - 256.5)^2 + (y - 256.5)^2\right]} - \sqrt{\left[(platpix)^2 + (platpix)^2\right]}\right\}\middle/ skirtpix\right) + 1\right]\middle/ 2$$

where *platpix* is the width of the central plateau of the envelope and was equal to *cycpix* times the number of nominal stimulus cycles (i.e., times the x-axis in the data figures) and *skirtpix* is the width of the blur skirt around the plateau and was equal to 2.