Katholieke Universiteit Leuven

Faculteit Psychologie en
Pedagogische Wetenschappen

Departement Psychologie



Doctoraatsprojecten

Early stages in spatial vision	
Goris, Robbe	26 juni 2007
Promotor: Prof. Dr. J. Wagemans	

Projectcommissie 26 juni 2007 (Psychologie)

Early Stages in Spatial Vision: Data and Models

Doctoraatsproject
R. Goris

Robbe Goris
Promotor: Prof. Dr. Johan Wagemans
Copromotor: Prof. Dr. Felix Wichmann
Laboratorium voor Experimentele Psychologie
April 2007

Abstract

Early Stages in Spatial Vision: Data and Models.

Robbe Goris, Doctoral project, April 2007
Laboratory of Experimental Psychology, University of Leuven
Supervisor: Prof. Dr. Johan Wagemans
Cosupervisor: Prof. Dr. Felix Wichmann

Despite the ease with which we perceive, it is not clear how the distribution of light across the visual field is analyzed by eye and brain to give rise to our meaningful percepts. One approach to study visual perception is to focus on the very first stages of information processing. Without precise knowledge of the computational processes that take place in early vision, it is difficult, if not impossible, to understand the more complex later stages in detail. Much evidence has accumulated that suggests that the early visual system is comprised of spatial frequency-selective channels. It is generally assumed that in contrast detection circumstances, these channels behave linearly. This assumption has been challenged by one study that reported contrast detection facilitation in the presence of weak levels of luminance noise (Blackwell, 1998). Improved information transmission in the presence of certain levels of externally added noise, i.e. stochastic resonance, is the signature of nonlinear information processing systems.

Contrast detection facilitation in the presence of weak levels of 2-D, white luminance noise is explored in Part I. Comparison of contrast thresholds measured in two different tasks, namely signal detection in Experiment 1a and orientation discrimination in Experiment 1b, but in otherwise identical circumstances, demonstrates that the origin of this facilitation effect is not to be found in a higher level task strategy. Minimization of spatial and temporal uncertainty about signal occurrence and the provision of feedback in Experiment 2 suggest that the visibility of low contrast signals is truly enhanced in the presence of weak noise levels. It is demonstrated that these data are consistent with a model of our visual system wherein a linear filter is followed by a nonlinear post-filter stage.

The implication of this finding for critical band masking, a technique to infer filter shape from detection in noise data is explored in Part II. In Experiment 3, by making use of low-pass filtered noise, i.e. noise from which all high frequency components have been removed, it is shown that the noise must not be white for detection facilitation to occur. Activation of the signal processing channel by the noise is sufficient. Experiment 4 demonstrates empirically that the power spectral density of the noise does matter in the critical band masking paradigm: Different noise levels lead to different results. A novel way to infer filter shape from these data is suggested.

Sinusoidal contrast discrimination in the presence of luminance noise is explored in Part III. Experiment 5 demonstrates that the well-known dipper-shaped threshold-vs.-contrast function (TvC function) changes somewhat in the presence of weak and strong noise levels. To be able to consider the effects of noise on contrast discrimination, irrespective of performance level, a computational contrast perception model was fitted to the data. Analysis of model deviances reveals that addition of weak noise diminishes the facilitating effect of a low contrast masker. Addition of strong noise causes a small rightward shift of the TvC function on double logarithmic coordinates. Whatever mechanism underlies contrast discrimination, future contrast perception models should aim to be able to explain these findings.

Recent work has questioned the validity of the current set of models of human spatial vision (Henning & Wichmann, 2007). The framework of a possible new model which might overcome some of the modelling problems is presented in Part IV. The central idea is to combine V1 neuron-like spatial frequency filters with a population output decision rule. Via simulations it is shown that this results in a model which reconciles constant tuning of the building blocks, as has been postulated often for cortical cells, with dynamic tuning properties of psychophysical channels.

Acknowledgements

First and foremost I want to thank Peter Zaenen, who is as a supervisor to me. Peter introduced me to psychophysics and low-level vision during my seven month internship at the Laboratory of Experimental Psychology in 2004-2005. We have been collaborating since that time, and the first two parts of this text largely stem from that joint work. I also wish to express my gratitude to Johan Wagemans and Felix Wichmann for their support and supervision. In addition, many thanks go to Hans Op de Beeck, Bruce Henning and Martin Lages for inspiring conversations and insightful comments. Most of all, however, I wish to thank Els Eerdekens, for everything.

Introduction

Object recognition and scene interpretation are easy activities for us. However, so far it is not clear how the distribution of light across the visual field is analyzed by eye and brain to give rise to our meaningful percepts. One approach to understand the essence of this process is to abstract away all factors that are not crucial for perception. A lot of factors that contribute to our daily visual experience can be omitted without impairing the ability to derive information from a pattern of light. These include motion, colour and the possibility to make use of both eyes. Removing relative differences between darker and lighter monochrome image regions, however, greatly damages our percept. One may thus argue that visual perception is primarily driven by the analysis of spatial variations in luminance, i.e. contrast. A first step towards understanding how we succeed in detecting and identifying objects and interpreting scenes is then to study how our visual system analyses spatial contrast.

Part I – Pattern Detection: Data and Models

Because of the importance of contrast, vision scientists have always had a great interest in the study of spatial luminance variation. Blackwell (1946), for instance, gathered 2 million measurements to determine the luminance increment threshold of the normal human observer under a variety of experimental conditions. But for long, a clear model specifying how the visual system processes spatial contrast was missing. Two crucial discoveries initiated the formulation of such a model. The first was the description of parametrically ordered neurons with specialized receptive fields in primary visual cortex (V1) by Hubel and Wiesel (1962). The second was the successful application of Fourier analysis to pattern detection by Campbell and Robson (1968). Back in 1956, Schade was the first to measure the human contrast sensitivity function (CSF, i.e. the inverse of the contrast detection threshold for single frequency sine wave gratings as a function of spatial frequency) and its typical inverted U-shape (see Figure 1). He suggested this shape was mainly caused by optical factors (i.e. light diffraction and modulation transfer). Later studies disentangled effects of the human eye and nervous system on the CSF (e.g. Banks, Geisler & Bennett, 1987; Campbell & Gubisch, 1966; Campbell & Green, 1965). The need for higher contrasts at lower spatial frequencies seems not to be a consequence of optical limits, but is instead largely driven by the nervous system. Fourier theory shows that all 2-D images can be considered as a sum of sine wave gratings (i.e. components), each with its own spatial frequency, phase, orientation and amplitude. By modelling the early visual system as a number of independent peak-detector mechanisms, each preceded by a narrowly tuned linear spatial-frequency filter, detection thresholds for any image can be predicted from the CSF (Campbell & Robson, 1968). This prediction will only be correct if the visual system processes contrast approximately linear, at least around threshold. This means that the response to a sum of two components equals the sum of responses to each component. Threshold measurements for a number of periodic gratings revealed a clear relationship between component detectability and overall detectability (Campbell & Robson, 1968). During the following years, the Campbell and Robson suggestion –independent, narrowly tuned linear spatial frequency filters– was tested in several studies. Blakemore and Campbell (1969) demonstrated that prolonged viewing of a high contrast grating (i.e. adaptation) leads to an elevation of detection thresholds for gratings of identical orientation and similar spatial frequency. They reported threshold elevations limited to a range of frequencies centred

on the adaptation frequency, and with a bandwidth of approximately one octave at half amplitude. By making use of estimates of the line-spread function of the human eye and modelling the early processing stages as a linear filter, Campbell, Carpenter and Levinson (1969) predicted CSFs for three aperiodic patterns. The empirical measurements agreed remarkably well with predictions. Campbell et al. (1969) concluded that *"[...] there is no reason to believe that a linear theory of threshold visibility would have to be abandoned"*. Experiments on the detectability of complex gratings (e.g. Graham & Nachmias, 1971), on simultaneous noise masking (e.g. Carter & Henning, 1971), and on pattern masking (e.g. Stromeyer & Klein, 1974) provided further evidence for the multiple channel model as suggested by Campbell and Robson (1968). One could speculate that the neural substrate of these channels is to be found in primary visual cortex (DeValois & DeValois, 1988). About twenty years after its initial formulation, two review works on the topic (DeValois & DeValois, 1988; Graham, 1989) concluded that all detection threshold studies were consistent with a model of our visual system in which low-level multiple channels with different tuning properties along different dimensions (e.g. spatial frequency and orientation) act in parallel and perform a local spatial frequency analysis. Strict linearity and independence are usually assumed as well, but small violations would not alter this conclusion dramatically. At higher contrasts, results are more ambiguous. Legge and Foley (1980) demonstrated that in simple contrast discrimination tasks the visual system behaves as a linear filter, followed by a nonlinear transducer. However, results that are hard to reconcile with linear, independent spatial-frequency filters have been reported as well at higher contrasts (e.g. Derrington & Henning, 1989; Henning, Hertz & Broadbent, 1975).

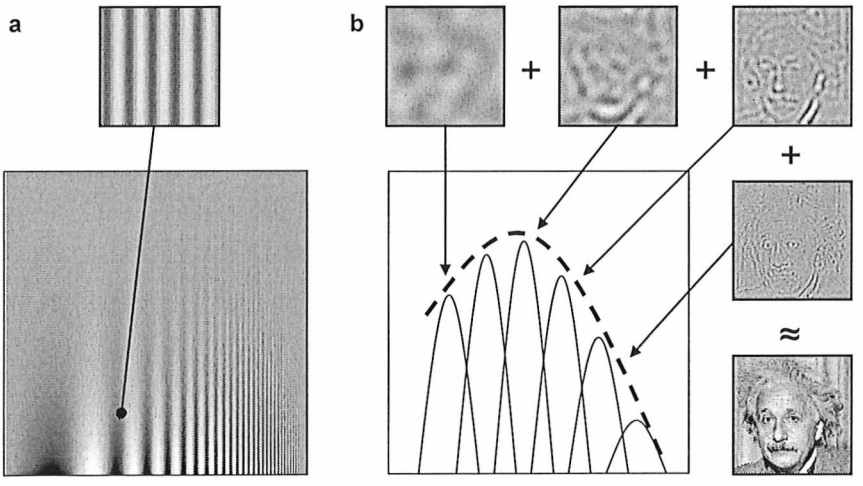


Figure 1
a. An easy way to illustrate the CSF. On the X-axis, spatial frequency rises logarithmically. On the Y-axis, contrast decreases logarithmically. The higher the black and white bars appear the more sensitive your visual system is for that specific frequency. **b.** Sensitivity and filter gain as a function of spatial frequency. The dashed line represents a CSF. Beneath this CSF, six hypothetical spatial frequency filters are drawn. The response for some of the filters is illustrated. A linear combination of these responses could approximate the input image.

Blackwell (1998) was the first, and so far only, to report that adding weak levels of white luminance noise to a contrast detection task leads to a decrease in thresholds (i.e. sensitivity increases). Such facilitation as a function of noise has been labelled stochastic resonance (Wiesenfeld & Moss, 1995). At higher noise levels, thresholds increase (i.e. sensitivity decreases). This latter effect has been described in many studies on contrast detection in luminance noise (e.g. Henning, Hertz & Hinton,

1981; Legge, Kersten & Burgess, 1981; Nagaraja, 1964; Pelli, 1985; Pelli & Farell, 1999; Solomon & Pelli, 1994) and is exactly what one would expect if the system under consideration processes information in a linear way. Stochastic resonance, on the other hand, is the signature of nonlinear information processing (see Figure 2). Its occurrence has been reported in a large number of biological (e.g. Douglass, Wilkens, Pantazelou & Moss, 1993; Levin & Miller, 1996; Russell, Wilkens & Moss, 1999) and man-made systems (e.g. Wiesenfeld & Moss, 1995).

In sum, Blackwell (1998) reported evidence that contrast detection in humans displays the signature of nonlinear information processing, whereas all other studies that have made use of external noise conclude that the system behaves linearly. There are two possible explanations for this deviance. The first is that Blackwell (1998) has got it wrong. Observers always had to discriminate between a signal and a blank, which differ on many dimensions (e.g., the signal always contains more contrast energy than the blank). This implies that it cannot be excluded that a higher level discrimination strategy has lead to this result (e.g., energy discrimination instead of contrast detection). The second is that Blackwell (1998) is right. Birdsall's theorem states that a (monotonous) nonlinear transformation that occurs after a relatively strong noise injection has no psychophysically measurable effect if the detection mechanism is only sensitive to order, as assumed in e.g. signal detection theory (Green & Swets, 1966; Pelli, 1991). One implication of this theorem is that strong external noise may effectively linearize a nonlinear system. Since all other studies have made use of relatively strong noise levels, the possibility that the region of facilitation simply has been missed, cannot be excluded. We ran two experiments in which we explored detection facilitation using weak external noise.

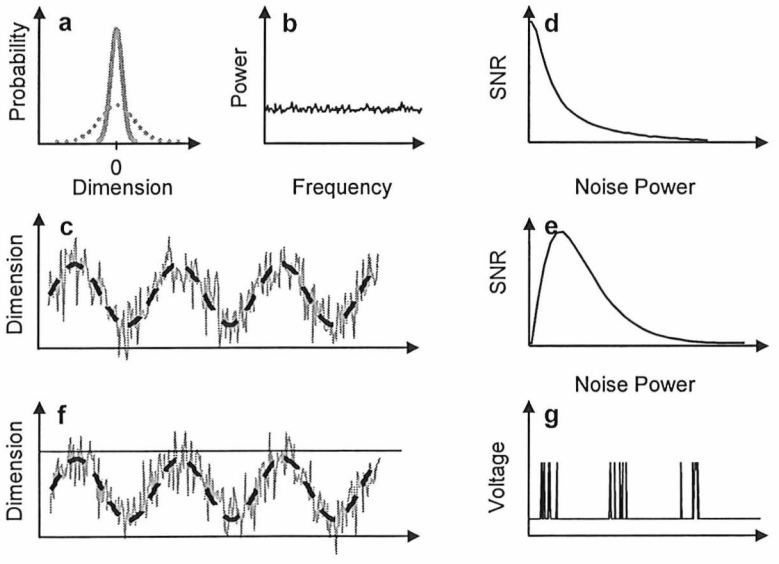


Figure 2

a. When white noise is added to a signal, noise values are sampled from a Gaussian distribution with mean zero. The spread of this distribution determines the strength (power) of the noise. Dimensions of noise and signal must be identical for the noise to 'stick' to the signal. **b.** In white noise, all frequencies are equally represented. **c.** A noiseless (dashed black line) and noisy (thin grey line) signal. **d.** The signal-to-noise ratio (SNR) at the output of a linear system will always decrease monotonically as a function of input noise. **e.** In nonlinear systems, stochastic resonance can occur. This means that the SNR at the output has a local maximum as a function of input noise. **f.** The simplest example of a nonlinear system in which stochastic resonance can occur is a hard threshold system. Suppose this system generates a spike (see **g.**) each time the input signal crosses the hard threshold (full black line). Without added noise, there would be no information transfer in this case because the signal is subthreshold. With added noise of appropriate power, there is some information transfer (see g.): The output *resonates* with the input signal thanks to the *stochastic* component added to the signal.

Experiment 1

To investigate whether noise facilitation is due to a nonlinear low-level contrast detection mechanism, or rather to a more high-level strategy like contrast energy discrimination, detection thresholds as a function of noise were measured in two different ways (see Appendix). In Experiment 1a, observers had to discriminate between a Gabor stimulus and a blank. In Experiment 1b, observers had to discriminate between two Gabor stimuli that were identical on all dimensions except for orientation (-45° and 45°). Both methods lead to very similar threshold estimates (Solomon & Pelli, 1994). If facilitation as a function of low noise levels is caused by nonlinear processing of contrast at threshold, both measurements should lead to a threshold-vs.-noise curve that has a local minimum at a nonzero level of noise. If, on the other hand, these data reflect a more high-level strategy, this should not be the case for the data of Experiment 1b. Results are summarized in Figure 3. Thresholds were first normalised for each participant by the noiseless detection threshold. This way, the noiseless detection threshold equals one and all other thresholds are expressed as products of this detection threshold. These normalised thresholds were averaged over observers and multiplied by the average noiseless detection threshold of all observers, to express the data on the original scale of measurement. Both experiments show facilitation as a function of low noise levels. In both cases optimal noise reduces the 69% correct threshold –69% correct in this type of Yes/No task corresponds to a *d'* of 1– by a factor of about 1.5 (1.56, with the 95% confidence interval ranging from 1.45 till 1.69, and 1.46, with the 95% confidence interval ranging from 1.38 till 1.53, respectively). This effect is neither caused by response bias (see Figure 3c, d), nor by monitor nonlinearities (see Figure 3e, f and Appendix).

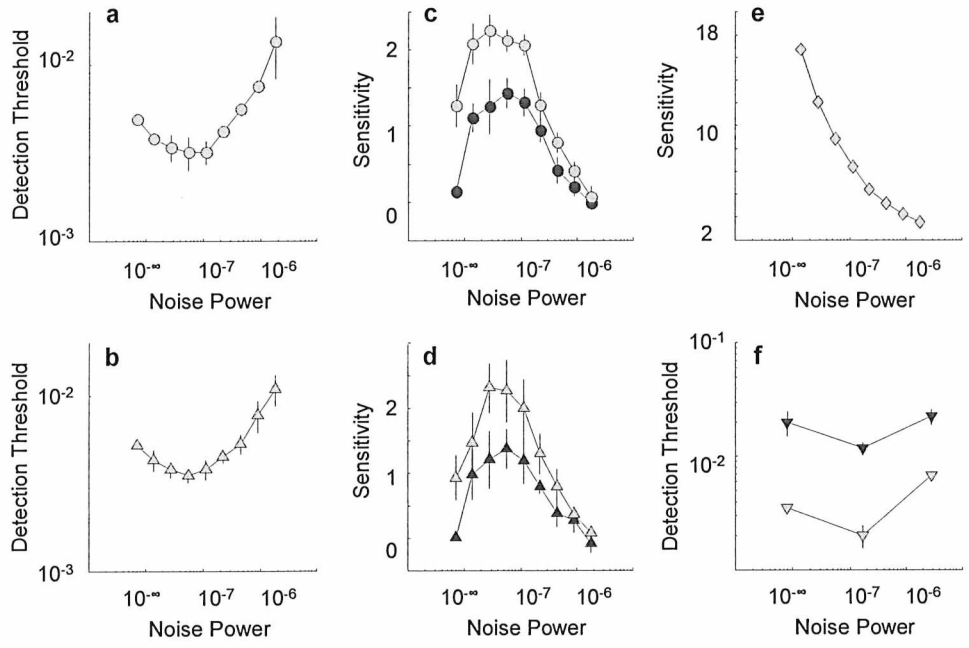


Figure 3

All thresholds are expressed in Michelson contrast. **a.** Average contrast detection thresholds as a function of noise power spectral density for Experiment 1a. **b.** Average contrast detection thresholds as a function of noise power spectral density for Experiment 1b. **c.** Average *d'* as a function of noise power for two different contrasts (.004 and .005) in Experiment 1a. **d.** Average *d'* as a function of noise power for two different contrasts (.004 and .005) in Experiment 1b. **e.** *d'* as a function of noise power for an ideal observer (see Appendix) for one contrast. Without added noise, *d'* is infinite for an ideal observer. **f.** Average contrast detection thresholds as a function of noise power for the control Experiment (see Appendix). Black triangles: fundamental frequency of 13.9 c/deg. Grey triangles: fundamental frequency of 2.8 c/deg.

Experiment 2

On some occasions, the human contrast sensitive system has been modelled with uncertainty models (see e.g. Pelli, 1985). The basic idea of this approach is that a detector monitors many channels. Only few of these are relevant to the presence of the signal. The detector usually bases its decision ('signal present' or 'absent') on the 'loudest' channel. The number of irrelevant channels that are monitored defines the uncertainty of this detector. Although the dimensions of uncertainty need not be specified, usually it is assumed that these are related to space and time of signal occurrence. Blackwell (1998) suggested that optimal noise facilitates detection because it reduces uncertainty about the spatial and temporal location of the signal. This suggestion was tested in Experiment 2 (see Appendix). As in Experiment 1b, observers had to discriminate between two orientations. Spatial and temporal uncertainty was minimized by a high contrast square surrounding the stimuli. Feedback was given to induce maximal performance. If the main contribution of optimal noise is to reduce uncertainty, facilitation effects should now disappear. Data were normalised, averaged over observers and rescaled as explained for Experiment 1. Results are summarized in Figure 4a.

Again, thresholds vary non-monotonically as a function of external noise. As in Experiment 1, optimal noise reduces the 75% correct threshold with a factor of about 1.5 (1.45, with the 95% confidence interval ranging from 1.28 till 1.66, respectively). 75% correct in a 2AFC-task corresponds to a *d'* of 1 and thus to 69% correct in a Yes/No task. The difference in performance (75% vs. 69%) is due to the additional information (2 images per trial vs. 1 image per trial) the observer has access to in a 2AFC-task (e.g. Wickens, 2002). The slope of the rising part of the curve is also comparable to the data of Experiment 1 (slightly lower than 0.5 on double logarithmic coordinates). The optimal noise level is about 3.5 times higher than in Experiment 1, probably due to the smaller pixel and stimulus size and shorter presentation time (see Appendix). These experimental variables were changed in comparison to Experiment 1 in order to raise detection thresholds: At higher contrasts, both signal and noise representation is better due to the larger range of luminance values. Note that the noiseless detection threshold in Experiment 2 is 1 log unit higher than in Experiment 1. The results of Experiment 2 suggest that noise enhances the visibility of the stimulus, rather than reducing temporal and spatial uncertainty.

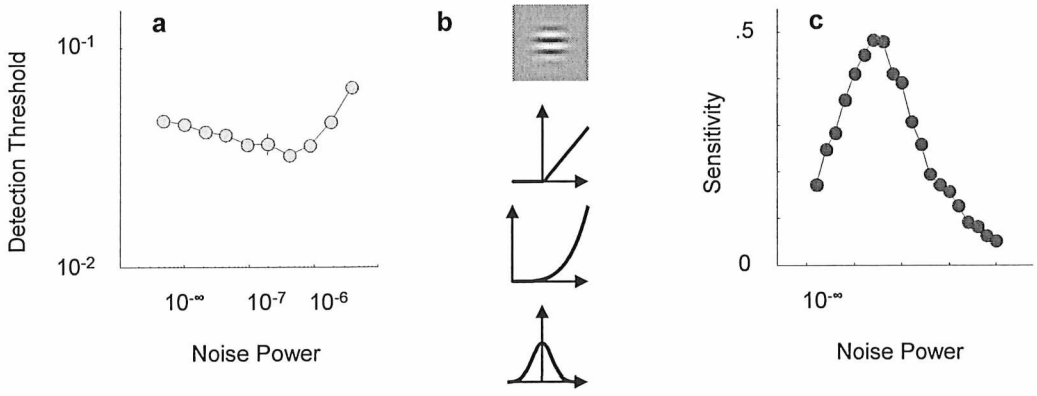


Figure 4

a. Average contrast detection thresholds (expressed in Michelson contrast) as a function of noise power spectral density for Experiment 2. **b.** The model used in the simulation had four stages: (1) a linear spatial frequency filter generating a 1-D response to 2-D input stimuli (Gabors and blanks), (2) half-wave rectification of this response, (3) sending it through a power-law with exponent 4 and (4) addition of late internal noise. **c.** Simulated *d'* as a function of noise power for one contrast value (compare to Figure 3c and d).

Discussion

Adding weak levels of white luminance noise improves human contrast detection performance. This effect is not caused by a more high-level strategy like energy discrimination. Also, it is not a consequence of spatial and temporal uncertainty reducing properties of the noise. In contrast, it seems to be a consequence of a low-level nonlinear contrast detection mechanism. Patel and Kosko (2005) have shown that several popular models of spiking retinal neurons exhibit stochastic resonance. In the next part, it will be demonstrated that integrating a linear filter with three generally accepted properties of cortical contrast sensitive neurons, as is often done in contrast perception models (see e.g. Legge & Foley, 1980), is sufficient to simulate results comparable to our behavioural effects.

Simulation 1

The goal of this simulation is to demonstrate that there is no need to introduce any new principle to obtain stochastic resonance-like behaviour. This is a to-be-expected phenomenon in Legge and Foley (1980) -like models of contrast perception. The first stage of the model consists of a linear spatial frequency filter, whose output is proportional to contrast for a noiseless sine wave grating with a spatial frequency corresponding to the peak sensitivity of the filter. Because cells cannot have a negative firing rate and linear filters do give negative responses, models that describe cell firing rate as a function of contrast typically include a (half- or full wave) rectification mechanism after this linear filter stage (e.g. Carandini et al., 2005; Heeger, 1992a,b). As other contrast perception models have done (see e.g. Foley, 1994), we shall adopt such a mechanism for our second stage. Studies on contrast detection have reported a positively accelerating *d'* as a function of grating contrast (see e.g. Foley & Legge, 1981). A *nonlinear transducer*, typically a power-law nonlinearity relating the output of the (rectified) filter's response to the decision variable has been suggested as explanation (see e.g. Legge & Foley, 1980). Again, this kind of input-output function has been postulated frequently for simple cells (see e.g. Albrecht & Geisler, 1991; Carandini, Heeger & Movshon, 1997; Miller & Troyer, 2002). This nonlinear transducer is our third stage. Since noise is inevitable in any biological system, addition of signal-independent noise is our fourth and last stage (see e.g. Legge & Foley, 1980). At this point, the simulated Gaussian distributions of the responses to the stimuli allow us to calculate *d'*-values. We simulated results for this model in a detection-in-noise experiment. As can be seen in **Figure 4b,** the combination of these four properties (linear filter, (half-wave) rectification, nonlinear transduction and additive noise) is sufficient to simulate stochastic resonance behaviour.

Conclusion

Ever since the famous study by Legge and Foley (1980), the idea that the visual system behaves as a linear filter, followed by a nonlinear transducer, additive noise and a decision stage in simple contrast discrimination tasks has been accepted. Usually, it is assumed that the nonlinearity can be safely ignored at detection threshold level. However, in this kind of model, stochastic resonance is a to-be-expected phenomenon. In other words, this kind of model predicts measurable effects of the nonlinearity at detection threshold level. This is exactly what we find. It is not the case that the linear models work because the nonlinearity has no effect at low contrasts. It does have an effect. There must be another reason why detection data often follow linear predictions. In the case of detection in (strong) noise, this reason could be the linearizing effects of early noise, as explained in Birdsall's theorem.

Part II – Inferring the Filters: Data and Methods

Since the very first suggestion of the existence of spatial frequency filters (Campbell & Robson, 1968), vision scientists have tried to infer the response characteristics of these filters. Despite the efforts to measure them, little is known about these hypothetical channels. A crucial aspect, on which all authors agree, is that they have different tuning properties along different dimensions (e.g. spatial frequency and orientation). But a lot of things are left unclear, for instance their rigidity or adaptiveness. Several psychophysical methods have been suggested to infer the shape of the filters mediating visual tasks. These include pattern adaptation (e.g. Blakemore & Campbell, 1969), subthreshold summation (e.g. Graham & Nachmias, 1971), pattern masking (e.g. Legge & Foley, 1980; Stromeyer & Klein, 1974) and critical band masking (e.g. Henning, Herz & Hinton, 1981). To be able to infer filter shape, all these methods must rely on assumptions. At the very least, one must specify how the filter processes contrast and how the resulting output is related to the psychophysically measured variables. So maybe it should not come as a surprise that these methods lead to very different estimates. It is not easy to interpret these differences. One should be aware that all filter estimates strongly rely on numerous assumptions, and that seemingly small differences in these assumptions may underlie severe differences in the inferred filters. This being said, we shall focus on the method of critical band masking. This is a relatively popular method (see e.g. Henning, 1988; Henning, Herz & Hinton, 1981; Majaj, Pelli, Kurshan & Palomares, 2002; Oruç, Landy & Pelli, 2006; Solomon, 2000; Solomon & Pelli, 1994; Talgar, Pelli & Carrasco, 2004) which makes use of detection in noise and is sometimes claimed to rely on only one assumption: The linearity assumption (more specifically; that thresholds are linearly related to the total noise power passed by the linear filter, see e.g. Solomon & Pelli, 1994). The logic behind critical band masking is illustrated in Figure 5a. Let us assume that the detectability of a signal solely depends on the signal-to-noise ratio at the output of a linear filter. This implies that changing the total amount of noise in the filter changes detectability of a constant signal processed by the same filter. In such a mechanism it is possible to derive the shape of the filter by manipulating characteristics of the noise on the dimension under consideration (in this case spatial frequency, see Figure 5a). However, this assumes that thresholds are linearly related to filtered noise power. It has been claimed that the validity of this assumption can be tested. If this assumption holds, the threshold elevation produced by a sum of two noises should equal the sum of threshold elevations produced by each noise alone (i.e. noise additivity). Applications of critical band masking in the context of spatial frequency channels, together with noise additivity, have been reported frequently. However, our previous results show that in white noise, noise additivity does not hold at all noise powers. If the reported detection facilitation is a consequence of a nonlinear post-filter stage, the spatial frequency characteristics of the noise should not matter. Detection facilitation should occur in band limited noise as well. This prediction will be tested in Experiment 3.

Experiment 3

In critical band masking experiments, the frequency characteristics of the luminance noise are manipulated. Typically, low- or high-pass noise is used. This means that no frequencies above or below a certain cut-off frequency are represented in the noise (see Figure 5a and b). If the stochastic resonance reported in Experiment 1 and 2 is a consequence of a nonlinear post-filter stage, similar threshold-vs.-noise curves should be obtained with non-white noise. Representation of the signal frequency in the noise is not critical, activation of the signal processing channel by the noise should be

sufficient. Low-pass noise with a cut-off frequency at 4 c/deg, for instance, will activate a channel with peak sensitivity at 8 c/deg and a bandwidth of 2 octaves at half amplitude. Because this noise invades this channel only marginally, the activation will be relatively small in comparison to the activation of the same channel caused by white noise of similar power spectral density as the low-pass noise. Consequently, one would expect that optimal low-pass noise will have a higher power spectral density than optimal white noise. Both predictions were tested in Experiment 3 by measuring detection thresholds as a function of low-pass noise power with a cut-off frequency slightly more than 1 octave below the signal frequency (see Appendix). Bandwidth estimates between 1 and 2 octaves at half amplitude have been reported frequently (see e.g. Majaj et al., 2002). Therefore, it is to be expected that our noise invades the channel only marginally. Hence, facilitation effects should occur at relatively high noise powers. Data were normalised, averaged over observers and rescaled as explained for Experiment 1. Results are summarized in Figure 5c. Facilitation as a function of noise power does occur in low-pass filtered noise: Maximal 75 % correct threshold reduction is about a factor of 1.85 (1.86, with the 95% confidence interval ranging from 1.40 till 14.86, respectively). This massive confidence interval indicates that it might be wise to rerun the experiment to find more reliable improvement estimates. Second, facilitation did occur at a higher noise power (2 log units higher) than in white noise. These results are consistent with the hypothesis that contrast is processed by spatial frequency channels of limited bandwidth, followed by a nonlinear post-filter stage.

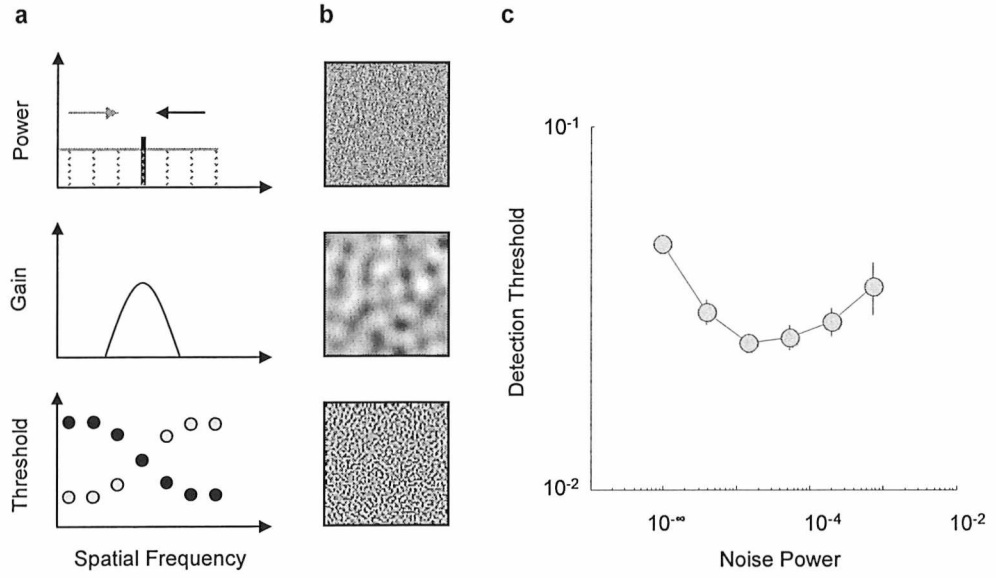


Figure 5
a. Critical band masking. **Above**: noise power as a function of spatial frequency. In white noise, all frequencies are represented. In low-pass noise, all frequencies above a certain cut-off frequency have been removed. In high-pass noise, all frequencies below a certain cut-off frequency have been removed. The thick black line represents the signal to be detected. The dashed grey lines represent several cut-off frequencies. **Middle**: filter gain as a function of spatial frequency. This filter is optimally tuned to detect the signal, because its centre frequency equals the signal frequency. **Below**: detection thresholds as a function of cut-off frequency. In the critical band masking paradigm, detection thresholds are measured as a function of cut-off frequency of low- (grey circles) and high-pass (black circles) noise. If the filter is linear and no nonlinear transformations occur before the decision mechanism, the threshold elevation caused by addition of a band of frequencies is directly related to the gain of the filter for that frequency band. Therefore, the shape of the filter can be derived by taking the first derivative of the threshold vs. cut-off frequency curve. **b.** Examples of 2-D white noise (above), low-pass noise (middle) and high-pass noise (below) stimuli. **c.** Average contrast detection thresholds (expressed in Michelson contrast) as a function of noise power spectral density for Experiment 3.

What is the relevance of these results to critical band masking? The linearity assumption implies that the power of the noise (in this case the height of the frequency spectrum, not the total filtered noise) does not influence derived filter shape. That is the beauty of the technique. Running the same experiment at two different noise levels would lead to two rescaled versions of the same threshold vs. cut-off frequency curves. Their normalised derivatives (i.e. the filter estimate) would be identical. Since our results suggest that noise additivity does not hold at all noise powers, however, we must predict that noise power does influence estimated channel shape: Running the same experiment at different noise levels might lead to very different threshold-vs.-cut-off frequency curves. This prediction was tested in Experiment 4.

Experiment 4

The shape of the threshold vs. white noise functions reported in Part I (see Figures 3a and b and 4a) suggests there are roughly three different types of threshold vs. cut-off frequency curves. Since the most extreme cut-off value equals white noise, we can distinguish (white) noise levels that facilitate, have 'no effect' and that inhibit detection. Running the critical band masking paradigm at each of these noise levels, should lead to three different threshold vs. cut-off frequency curves: one that only decreases as a function of cut-off frequency, one that first decreases (when the total filtered noise power is 'optimal') and then rises till detection threshold level and one that might decrease initially, but then rises above detection threshold level[1]. In Experiment 4, we ran the critical band masking paradigm with low-pass noise at three different noise levels (see Appendix). It is important to realise that critical band masking studies usually make use of noise levels that are at least as high as the highest of our noise levels. Data were normalised, averaged over observers and rescaled as explained for Experiment 1. Results, summarized in Figure 6, confirmed our predictions. Maximal threshold reduction is about a factor of 1.4 this time (1.41, with the 95% confidence interval ranging from 1.25 till 1.53, respectively).

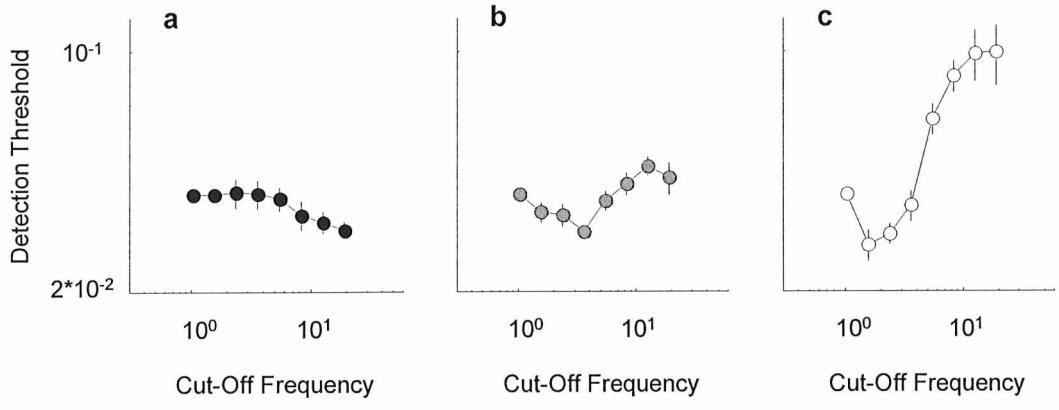


Figure 6
a. Average contrast detection thresholds (expressed in Michelson contrast) as a function of cut-off frequency for the lowest noise level used in Experiment 4. **b.** Average contrast detection thresholds as a function of cut-off frequency for the middle noise level used in Experiment 4. **c.** Average contrast detection thresholds as a function of cut-off frequency for the highest noise level used in Experiment 4.

[1] These predictions are formulated from the point of view of low-pass noise. For high-pass noise, predictions are identical if the curves are considered from right to left.

Discussion

In Part I, we concluded that the observed facilitation as a function of white noise is likely to be a consequence of a low-level nonlinear contrast detection mechanism, e.g. a linear filter followed by a nonlinear transducer. Because contrast is assumed to be processed within channels that have a certain bandwidth, the noise must not be white to observe facilitation and inhibition: Invasion of the signal processing channel by the noise is sufficient for these effects to occur. This prediction was confirmed in Experiment 3. This led us to predict that the power spectral density of the noise does matter in the critical band masking paradigm: Different noise levels should lead to different results. Again, this prediction was confirmed. These data must not be inconsistent with other applications of critical band masking. Usually, only noise levels that are at least as high as our highest noise level are considered. Furthermore, Oruç, Landy and Pelli (2006) have reported an 'unexplainable' initial threshold decrease, consistent with our current observations.

What do these results imply about the derivation of filter shape through critical band masking? The most straightforward answer is that Pelli and Solomon's (1994) assumption that noise additivity holds at all noise powers must be reconsidered. This assumption may hold when the noise is strong enough to inhibit contrast detection, but it does not at lower noise powers. There are two fundamental unknowns: the shape of the filter and the nonlinear transformations that occur. At least some specifications must be made to be able to derive information about one of them. In the next part, both the 'classical' and an alternative method to infer filter shape from these data will be considered.

Simulation 2

The goal of this analysis is to derive the spatial frequency response characteristics of the channel that is assumed to mediate detection of the sine wave grating in Experiment 4. All analyses will be run on the averaged data. First, we will try to mimic the analysis described by Solomon and Pelli (1994). For this purpose, we can only make use of the data gathered at the highest noise level in Experiment 4 (see Figure 6c). Otherwise the noise additivity assumption would be violated. Because this is also the case if we include the thresholds for the two lowest cut-off frequencies, these data points will be omitted for this analysis as well. It has been claimed that the first derivative of the sigmoid shaped energy-threshold-vs.-cut-off frequency function specifies the channel the observer uses to detect the signal. As can be seen in Figure 7b, the contrast energy channel derived this way is most sensitive somewhat above the signal frequency (8.25 c/deg). At half amplitude, bandwidth is about 1.7 octaves (8.25 – 0.88 and + 0.84 octave). This bandwidth is not inconsistent with other applications of critical band masking (see e.g. Majaj et al., 2002). In sum, if we only consider (most of) the data gathered at the highest noise level, we can apply the technique suggested by Solomon and Pelli (1994) to derive filter shape. This filter is off-frequency, but otherwise comparable to filters reported by others making use of similar techniques. It is important to realise that this filter is not unique. Henning et al. (1981) have shown that the exact shape of the threshold-vs.-cut-off frequency curve strongly depends on the contrast statistic one considers. To illustrate this dependency, a plot of the filter derived exactly the same way from the same data, but this time expressed in Michelson contrast, has been added (see Figure 7b). In sum, one needs to make assumptions about the contrast statistic of interest to our visual system as well, to be able to infer filter shape from these data. Note that noise additivity cannot hold for many contrast statistics at the same time, because some of them (e.g. Michelson contrast and contrast energy) are nonlinearly related to one another.

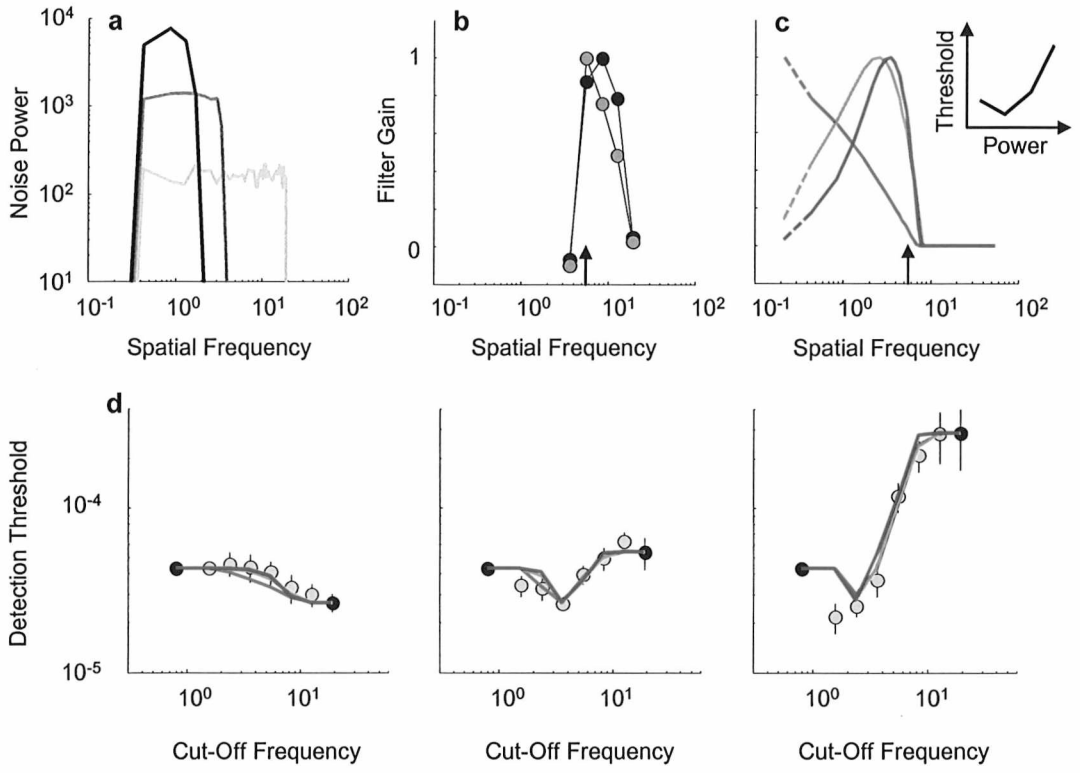


Figure 7
a. In Experiment 4, a 5.5 c/deg Gabor stimulus had to be detected in the presence of low-pass noise. Three different noise levels and seven different cut-off frequencies were used. Here, the average power spectra of the vertically oriented components of 100 *estimates of effective images* of three instances of noise are shown. **b.** Black filter: The normalised spatial frequency energy filter derived from (most of) the data gathered at the highest noise level in Experiment 4 by the method of Solomon and Pelli (1994). If one would assume that Michelson contrast is the contrast statistic of interest of our visual system, rather than contrast energy, this method would result in a slightly different filter (grey filter). **c.** Three spatial frequency filters derived from all the data by the method described in Simulation 2 (red line: rectification filter, green line: energy filter, blue line: double squaring filter). Inset: The (filter specific) nonlinear transformation function relating filter response-to-noise to detection threshold. **d.** The fit of these filters to the data of Experiment 4. The data are expressed in contrast energy. Note that prediction for the first and last data point must be correct.

So far, all our results are consistent with a contrast detection mechanism consisting of a linear spatial frequency filter followed by a nonlinear post-filter stage. As stated above, if we are to infer information about filter shape from the data of Experiment 4, we have to make some specifications about the nonlinear stage. I opted for a rather empirical approach that remains agnostic about the underlying cause of this nonlinearity.

First of all, we must assume that detection of the signal in Experiment 4 was mediated by a single spatial frequency filter which remains constant throughout the experiment. Otherwise, it would not make any sense to try to infer this filter. My point of departure is that it seems not unlikely that the empirically observed effects of external noise (i.e. facilitation and inhibition) are directly related to the strength of the response of the detection channel to luminance noise: Weak stimulation of the channel facilitates detection, while strong stimulation inhibits detection. This might explain why both addition of white (Experiment 1 and 2) and low-pass noise (Experiment 3) to a detection task give rise to a dipper shaped threshold-vs.-noise power curve. By specifying a (nonlinear) function relating this filter response-to-noise to detection threshold, the goodness-of-fit/badness-of-fit of any candidate filter to the data can be calculated. For each possible filter, the response to any noise condition can be

simulated. Because the input images are stochastic, the average filter response to 25 *estimates of the effective noise images* was used. To specify the nonlinear function, these simulated responses must be linked to the empirically measured (energy) thresholds for the same noise conditions. Four conditions of Experiment 4 were selected for this purpose: the noiseless condition and the three highest cut-off conditions. This way, the whole range of filter responses to the noise conditions of Experiment 4 was covered. Furthermore, both the dip and initial rise of the nonlinear function were approximated (see Figure 7c, inset). It will be noted that this procedure results in filter-specific nonlinear transformation functions. Once this filter-specific function has been specified, detection thresholds of this filter can be predicted for all conditions of Experiment 4. Because the noiseless detection threshold and the thresholds of the three highest cut-off conditions were used to specify the nonlinear transformation, predicted values for these 4 conditions are always correct (see Figure 7d). This is not the case for the 18 remaining data points. To obtain an optimal fitting filter, the *weighted sum of squared errors* was minimized as a function of filter characteristics (i.e. peak sensitivity, bandwidth and slope of both left- and right filter side). Finally, to consider the effect of contrast statistic, three different types of spatial frequency filters were used: a *Rectification filter* (each Fourier component is rectified before integration of all components takes place), an *Energy filter* (each component is squared before integration) and a *Double squaring filter* (each component is raised to the power of four before integration).

For each minimization routine, the same set of noise images was used. To account for the variability inherent to noise images, this analysis was repeated 10 times. For each type of filter, the filter estimates for these 10 sets of noise images were virtually identical. The fit of these filters (and nonlinear transformations) to the data of Experiment 4 was satisfying (see Figure 7d). The derived energy channel has peak sensitivity at 2.62 c/deg. At half amplitude, bandwidth equals about 3 octaves (2.62 – 2.29 and + 1.15 octave). At half amplitude, the channel is asymmetric. However, it is clear that these properties strongly depend on the contrast statistic of choice (see Figure 7c).

In sum, a new method is suggested. This time, all data points are included in an analysis that varies filter characteristics and makes use of a largely empirically determined nonlinear transformation to relate filter response-to-luminance-noise to measured thresholds. The weighted sum of squared errors was minimized as a function of filter characteristics. The resulting estimates were shown to be stable, despite the variability inherent to noise images. The fit was satisfying. As in the classical critical band masking paradigm, the specific filter characteristics strongly depend on the contrast statistic of choice.

Conclusion

Despite the success of the multiple channels model, little is known about these spatial frequency channels and their role in psychophysical tasks. Critical band masking is a popular approach to derive their frequency response characteristics. Within this approach, it is usually assumed that these filters process contrast linearly. However, by making use of relatively weak noise levels, we have demonstrated that noise power does matter within this paradigm. At the very least, we must conclude that the assumption of noise additivity does not hold at all noise powers. To be able to derive filter shape from band masking data, at least some assumptions must be made about the processing of contrast. An alternative method that assumes nonlinear contrast processing was suggested. Within this approach, the importance of the assumptions one makes about the contrast statistic of interest to the visual system was demonstrated as well.

Part III – Suprathreshold vision: Data and Models

Even without the notion of linear and independent channels, the multiple channels model still captures many aspects of early spatial vision at detection level. But to be truly useful, a model of spatial vision must explain how the visual system processes suprathreshold contrasts as well. One way to gain insight in the visual system's operation at suprathreshold contrasts is to study sinusoidal contrast discrimination (see e.g. Foley, 1994; Foley & Legge, 1981; Gorea & Sagi, 2001; Henning & Wichmann, 2007; Legge, 1981; Legge & Foley, 1980; Legge, Kersten & Burgess, 1981; Nachmias & Sansbury, 1974; Pelli, 1985; Wichmann, 1999). The main finding of these studies is the pedestal or dipper effect: A sinusoidal grating is much more detectable when added to a low-contrast masking or pedestal grating of identical spatial frequency, orientation, phase, and temporal duration. Once the contrast of the pedestal grating exceeds a certain value, discrimination thresholds rise in a Weber's law-like fashion (see Figure 8a). The main explanations that have been suggested for the pedestal effect are a nonlinear transducer (see e.g. Legge & Foley, 1980) and contrast gain control (for a more exhaustive review of suggestions, see e.g. Foley, 1994; Wichmann, 1999). These and other suggested explanations all try to model the dipper effect as arising from characteristics of a single spatial-frequency-tuned channel. Very recently, serious doubt has been casted on this assumption and thus on all these models. Henning and Wichmann (2007) demonstrated that the dipper effect persists if strong 1-D white noise is added to the discrimination task, but disappears almost completely if notched noise –noise from which a 1.5-octave band centred on the signal frequency had been removed– is added. Furthermore, the pedestal effect remains substantial if either low- or high pass noise is used. These data cannot be explained by any of the current models. In contrast, they suggest that the classical pedestal effect results from off-frequency looking: The observer bases his decisions on the responses of channels with peak sensitivities at spatial frequencies different from that of the signal and pedestal.
Despite the uncertainty about the mechanism underlying the pedestal effect, it is conceivable that a similar mechanism underlies stochastic resonance in contrast detection. Indeed, adding luminance noise to a signal grating not only results in stimulus variability, but also increases the contrast of the stimulus. On the other hand, the contrast increase due to the addition of 'optimal' noise is negligible in comparison to the contrast increase due to the addition of an 'optimal' pedestal. Furthermore, the pedestal effect strongly depends on the phase relation between pedestal and signal: Introducing a phase shift of 90° removes the dip (Wichmann, 1999; Yang & Makous, 1995). In white noise, all phases are represented. Depending on the assumptions one makes about the processing stage of the observer (phase-sensitive vs. phase-insensitive), this allows or prevents the pedestal-mechanism to play its role. In sum, two different non-linear mechanisms might give rise to stochastic resonance and the pedestal effect, independently of one another. Alternatively, the same mechanism, possibly off-frequency looking, could underlie both phenomena. Both options (one vs. two non-linear mechanisms) would lead to different contrast discrimination data patterns in the presence of luminance noise. Experiment 5 was designed to explore sinusoidal contrast discrimination in luminance noise.

Experiment 5

To investigate whether the same non-linear mechanism underlies the stochastic resonance observed in contrast detection and the pedestal effect seen in contrast discrimination, the pedestal effect was measured without externally added noise, in noise that facilitates detection and in noise that inhibits detection (see Appendix). It has been noted before that addition of (strong) white noise does not

change the shape of the TvC-function (i.e. Thresholds-vs.-Contrast function). The relative depth of the dip on logarithmic coordinates (i.e. the factor of improvement) sometimes diminishes, however (see e.g. Henning & Wichmann, 2007). This might simply be a consequence of the stimulus variability introduced by the noise: A noisy optimal pedestal will on average have the optimal contrast, but on many stimulus presentations a slightly higher or lower contrast, both leading to higher thresholds. If two different mechanisms underlie both effects, we would expect that the normalised TvC-curves are almost identical for the no noise and facilitating noise condition. If the introduced stimulus variability has any dip-reducing effect, this should be less severe for the curve measured in facilitating noise than for the one measured in inhibitory noise because the latter one introduces more variability. In other words, if two different mechanisms underlie both effects, the relative depth of the dip should be most pronounced in the no noise condition and at least as deep in the facilitating noise condition as in the inhibitory noise condition. Results are summarized in Figure 8a. 75% correct thresholds, averaged among participants, are plotted. Both the intended noise facilitation and inhibition are present in the detection conditions (the facilitation with a factor of about 1.31, with the 95% confidence interval ranging from 1.11 till 1.45, and the inhibition with a factor of about 1.47, with the 95% confidence interval ranging from 1.32 till 1.64; compare the most leftward grey and white to the black data point). Furthermore, it is clear that the shape of the TvC-function is preserved in both noise conditions. At the 75% correct threshold, the maximal pedestal-improvement displays a trend that is inconsistent with the 'two independent mechanisms' hypothesis (without added noise, the maximal improvement has a factor of about 3.32, with the 95% confidence interval ranging from 1.85 till 7.95; in weak noise this factor drops till 2.20, with the 95% confidence interval ranging from 1.57 till 3.30; and in strong noise it is 2.70, with the 95% confidence interval ranging from 1.87 till 4.04). However, the pedestal effect changes with the threshold level one considers (Bird, Henning & Wichmann, 2002; Wichmann, 1999; see Figure 8c). There is no reason but convention to prefer the 75% correct threshold over other performance levels. To be able to consider the effects of noise on contrast discrimination irrespective of performance level, a contrast perception model that expresses performance level as a function of pedestal- and increment contrast was fitted to the data (see Appendix).

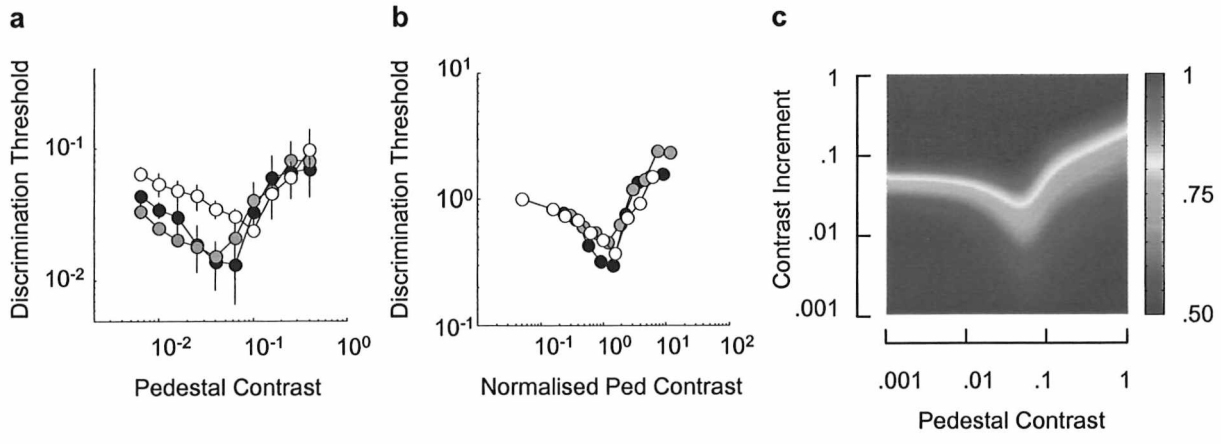


Figure 8
a. Average contrast discrimination thresholds (expressed in Michelson contrast) as a function of pedestal contrast for three noise levels (black = no noise, grey = facilitating noise, white = inhibitory noise). 75% correct thresholds are plotted. The most leftward data point denotes detection (pedestal contrast = 0). **b.** The same data, normalised by their respective detection thresholds. **c.** The pedestal effect at all performance levels, according to the model fit to all data (see text).

A Descriptive Contrast Perception Model

This model, introduced by Wichmann (1999), has many attractive properties, even when only considered for its descriptive value. The reported goodness-of-fit to very extensive datasets is impressive (Wichmann, 1999). Furthermore, all available data information is used by fitting performance level as a function of pedestal- and increment contrast. Finally, this model allows considering the pedestal effect at all performance levels. Under the null-hypothesis that addition of luminance noise has no effect on contrast discrimination, this model can be fitted to all data at once. To be able to do so, the noise data (both pedestal and threshold contrasts) must first be normalised by their respective 75% correct detection threshold (see Figure 8b) and then rescaled by the noiseless 75% correct detection threshold. This way, all data are expressed on the same scale (i.e. the no noise condition scale). Careful analysis of changes in parameter values and deviances per noise condition will then decide whether the null-hypothesis must or must not be rejected.

The model was fitted both to the data of the no noise condition only and to all (rescaled) data. If externally added luminance noise does not influence contrast discrimination, there should be no major change in parameter values. However, this is not the case (see Table 1). Two of the six parameters (η and κ, i.e. the exponents of the nonlinear transducer) undergo only a mild change, but the other four parameters increase or decrease by at least a factor three. Apparently, addition of noise has some effect on contrast discrimination. Analysis of the deviances (to the global fit) per noise condition should enable us to judge where these conditions differ from each other.

Parameters	α	β	η	κ	ζ	γ
Noiseless Data	$3.99e^{4}$	$6.71e^{-6}$	4.66	4.12	$2.29e^{3}$	0.021
All Data	$1.37e^{5}$	$5.35e^{-5}$	4.35	3.57	276.26	0.340
Ratio	0.29	0.13	1.07	1.15	8.30	0.06

Table 1
The parameter values that provide the best fit of the model for the no noise condition, for all conditions, and their ratios.

Results

The deviances were calculated by subtracting the observed performances from the predicted performances (see Figure 9). Therefore, positive values indicate an overestimation of performance and negative values an underestimation. Per noise condition, deviances were split in two groups: those stemming from the facilitating region of the pedestal effect and those stemming from the inhibitory region. The separation line (i.e. the deepest point of the dipper function) was determined by the model fit (approximately 4.5 % Michelson contrast). A t-test was used to judge whether any systematic bias was present, linear regression to judge whether any linear trend (i.e. a slope deviating from 0) could be found. Per test, the α-level (i.e. the criterion to judge whether an effect is significant) was 0.01, still leading to a considerable global risk for type I errors, due to the 12 tests in total (3 noise conditions * 2 data halves * 2 tests = 12 tests; $p(\text{type I error}) = 1 - (1 - \alpha)^{12} = 0.11$).

There was no systematic bias in the deviances of the no noise condition, nor any linear trend (see Figure 9, left column). This should not be taken as evidence for the null hypothesis (*"addition of luminance noise has no effect on contrast discrimination"*), but simply means that any effect that might be found in the reminder of the data can be expressed relative to the 'normal pedestal effect'. A

systematic bias was found in the facilitating region for the weak noise condition (see Figure 9, middle column): The improvement due to the presence of the pedestal is overestimated. On average, performance was overestimated by 7.6 %. In other words, the facilitating effect of low pedestal contrasts diminishes in the presence of noise that facilitates detection. There was little evidence for systematic bias in the inhibitory region for the weak noise condition (see Figure 9, middle column). On average, performance was underestimated by 2.75 %. There was no linear trend in the deviances of the weak noise condition. Inspection of Figure 9, right column shows that a systematic bias was found in the inhibitory region for the strong noise condition. On average, performance was overestimated by 5.13 %. Furthermore, a linear trend was clearly present in the facilitating region for the strong noise condition (slope = -3.90). Together, these observations suggest that the dipper function has shifted somewhat to the right on double logarithmic coordinates in the strong noise condition. This effect should not be overestimated (see Figure 8b).

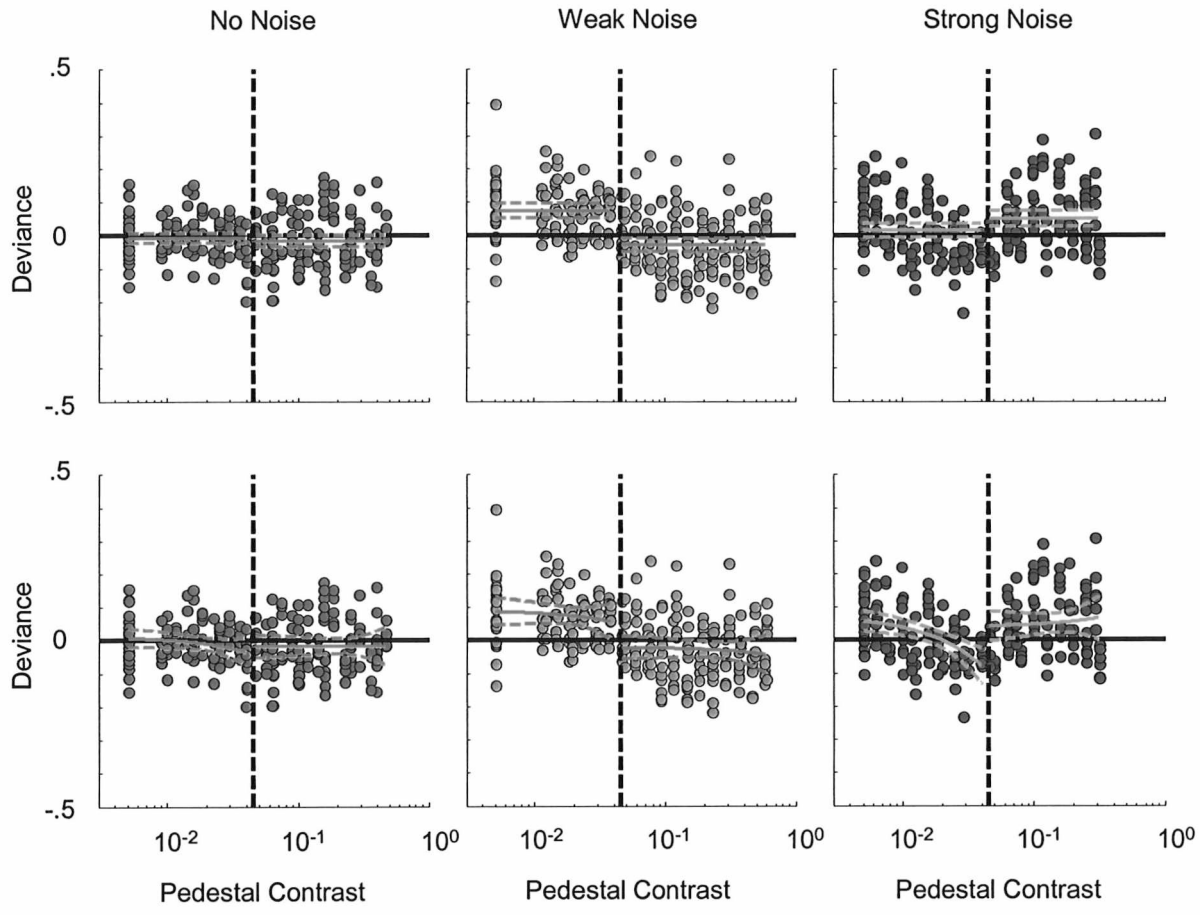


Figure 9
The deviances to the model fit (see **Figure 8c**) per noise condition. Upper row: t-statistic with the 99% confidence interval. Lower row: linear regression with the 99% confidence band.

Discussion and conclusion

To gain insight in the visual system's operation at suprathreshold contrasts, sinusoidal contrast discrimination has been studied extensively. The main finding of these studies is the pedestal or dipper effect. Recently, it has been suggested that off-frequency looking underlies this effect. Whatever

mechanism causes the pedestal effect, it is conceivable that this is the same mechanism that underlies stochastic resonance in contrast detection. On the other hand, it cannot be excluded that two independent mechanisms are at work. In the latter case, one would expect that the dipper function undergoes no major changes when measured in the presence of luminance noise. We explored contrast discrimination without, in weak and in strong luminance noise in Experiment 5. To be able to consider the effects of noise on contrast discrimination, irrespective of performance level, a contrast perception model that expresses performance level as a function of pedestal- and increment contrast was fitted to all (normalised and rescaled) data. Analysis of the deviances to this model fit revealed that the facilitating effect of low pedestal contrasts diminishes in the presence of noise that facilitates detection. This finding is inconsistent with the 'two independent mechanisms' hypothesis. Furthermore, these data suggest that the dipper function shifts somewhat to the right on double logarithmic coordinates in the presence of noise that inhibits detection. Whatever mechanism underlies contrast discrimination, future contrast perception models should aim to be able to explain these findings.

Part IV – Future work: Towards a new model of spatial vision

Ever since the breakthrough of mathematical psychology in the first half of the 20th century, the idea that human sciences should strive for precise, quantitative models has gained popularity. Not surprisingly, many computational models have been proposed in low-level vision during the last 40 years. And not without success: It is hard to find other examples in human sciences where data follow predictions that closely. However, this is not to say that I find the current set of models of spatial vision satisfying. Several proposed models have clear merits, but they all fail to explain non-trivial deviations from their predictions. A second problem lies in the wide variety of models that have been proposed to explain different aspects of the behaviour of what ultimately must be one system. And finally, considering the biological implementation, there is often an uncomfortable distance between psychophysical models on the one hand and physiological data on the other hand. In the first paragraph of Part IV, a very short overview of the main merits and failures of some of the most important models will be given. In the second paragraph, the framework of a possible new model, which aims to overcome the problems sketched above, will be proposed.

Models of spatial vision

The multiple channels model

As seen in Part I, the multiple channels model was initially proposed by Campbell and Robson (1968). The conceptualisation of the early visual system as a number of independent, linear spatial frequency filters of limited bandwidth offered a simple and plausible framework to explain detection of and discrimination among low-contrast periodic patterns. However, it is not really clear what the exact scope of the model is. Some studies which made use of suprathreshold patterns reported results consistent with the model (e.g. Carter & Henning, 1971; Stromeyer & Klein, 1974), while others casted serious doubt on the independence of the spatial-frequency channels (Albrecht & DeValois, 1981; Derrington & Henning, 1989; Henning, Hertz & Broadbent, 1975). It is hard to interpret these differences. Furthermore, it was concluded in Part I that the nonlinear transduction which is often assumed to follow the linear filter stage has an effect on detection thresholds.

Uncertainty models

The most popular among the uncertainty models is probably the one proposed by Pelli (1985). Uncertainty models conceptualise the early visual system as a detector that monitors both signal-relevant and signal-irrelevant channels. The decision mechanism usually consists of a maximum output rule (i.e. the detector bases its decision on the output of the loudest channel). This way, uncertainty models can explain some of the deviations from ideal observer performance typically found in human spatial vision. However, it has been noted on several occasions that uncertainty models fail to explain the pedestal effect in a satisfying way (e.g. Wichmann, 1999). Furthermore, these models predict monotonically rising thresholds if luminance noise is added to a detection task. This is inconsistent with Blackwell (1998) and our own results.

Divisive Inhibition/Contrast gain control models

These models go back to a very successful model of V1-cells, proposed by Heeger (1992a,b). In these models, each channel has both a relatively small band excitatory and a relatively broadband inhibitory input (e.g. Foley, 1994). This way, these models explain successfully how the pedestal effect changes if the orientation of mask and signal differ. An extended version of the model was shown to provide an excellent fit to contrast discrimination data (Wichmann, 1999). However, despite the clear link with physiology, the estimated parameter values do not match with the typical parameters values that fit to the contrast response functions of cortical cells (e.g. Geisler & Albrecht, 1995; Geisler & Albrecht, 1997). More important, these models fail to explain why the pedestal effect abolishes if notched noise is added to a contrast discrimination task (Henning & Wichmann, 2007).

Towards a new model

Building Blocks

It is generally agreed upon that the V1-cell model proposed by Heeger (1992a,b) captures many aspects of nonlinear cell behaviour (e.g. Carandini et al., 2005). Essentially, the (simple) cell output consists of a half-rectified linear spatiotemporal filter response, which is subject to divisive inhibition from all other neurons whose receptive fields cover the same part of the visual field (see Figure 10). Recently, Carandini, Heeger and Senn (2002) and Freeman et al. (2002) have suggested that the neurophysiological mechanism might be quite different from the one initially proposed by Heeger – they propose thalamocortical synaptic depression as an alternative for intracortical inhibition–, but the arithmetic remains more or less the same. Several studies have concluded that the response variance of cortical neurons is almost completely captured by simply rescaling the mean with a factor of about 1.5 (e.g. Geisler & Albrecht, 1995; Vogels, Spileers & Orban, 1989). This mechanism was introduced in our building blocks. One could argue that a collection of such '*Heeger-units*', with different tuning properties, is a very restrictive, but plausible model for primary visual cortex activity in a contrast discrimination task. Despite the mild parameter-variability typically found in cortical cells (Albrecht, Geisler & Crane, 2002; Geisler & Albrecht, 1995; Geisler & Albrecht, 1997), all the cells in these simulations give energy responses (i.e. the exponent of both the inhibitory and excitatory input equals two), for the sake of clarity and simplicity.

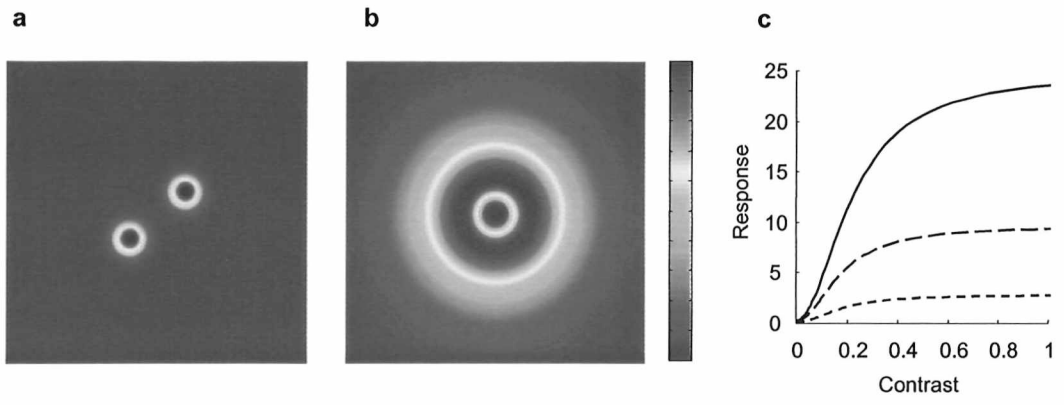


Figure 10
a. The excitatory input of a *Heeger-unit* in Fourier space on a polar plot (spatial frequency bandwidth at half amplitude = slightly less than 1 octave, orientation bandwidth = 30°). **b.** The inhibitory input of the same Heeger-unit (spatial frequency bandwidth = 2 octaves, orientation bandwidth = 360°). **c.** The mean responses of three units with different spatial frequency tuning to the same Gabor-stimulus. Note two nonlinear properties: response expansion at low contrasts, and contrast-set response saturation (gain control).

Decision Mechanism

It has been shown that (optimal) pooling of cell responses is an elegant and plausible way to relate single cell behaviour to psychophysical performance (e.g. Geisler & Albrecht, 1995). Introducing such a mechanism implies that decisions ('signal present' or 'signal absent') are based on a weighted average of the responses of many cells. The rule that is used to calculate the weights for each cell will determine many aspects of the behaviour of this model. Given that contrast discrimination might be a very unnatural task for the visual system –at neural level, contrast information is largely sacrificed in order to reach constant spatial frequency and orientation tuning–, and that each cell's response contains some, albeit little, information referring to the contrast of the stimulus, a signal-to-noise ratio statistic seems a possible combination rule. In these simulations, each cell's weight was determined by the ratio of the average response and the standard deviation of this response. A final important consideration is that stimulus variance inevitably enhances the variability of the cell responses.

Simulated Model Behaviour

In this model, the tuning of a psychophysical channel is not constant. For instance, the tuning properties of the different cells which underlie 'the channel' remain constant, irrespective of stimulus contrast, but the weight each cell gets, changes with stimulus contrast (see Figure 11a). It will be noted that the off-cells make a considerable contribution to the decision statistic (i.e. the weighted average response function). This is due to the fact that the response variance of each cell is almost completely signal-dependent. Therefore, the signal-to-noise ratio, and consequently the weight, of on- and off-tuned cells does not differ that much. In comparison to the response function of the most sensitive cell, the weighted average response function has (1) a very restricted dynamic range, (2) a much improved signal-to-noise ratio at each contrast level and (3) a more hard threshold-like character (see Figure 11b). In other words, were the decisions of the model based on the response function of the most sensitive cell instead of the weighted average response, contrast discrimination performance would be worse at all pedestal levels and the dip of the dipper function would be less pronounced (see Figure 11c).

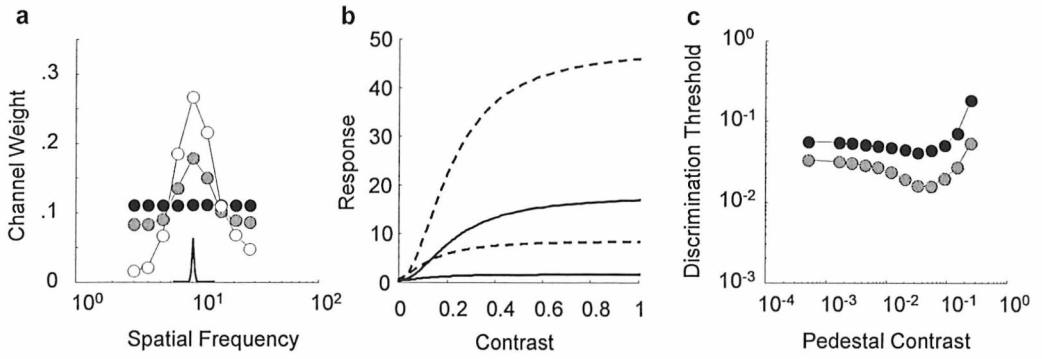


Figure 11
a. The distribution of weights for 9 cells with peak sensitivities geometrically scaled between 3 and 27 c/deg. A 9 c/deg Gabor stimulus was presented at three contrast levels: .1% (black), 1% (grey) and 10% (white). **b.** The means (upper two curves) and standard deviations (lower two curves) of the contrast response function of the most sensitive cell (dashed lines) and the weighted average (full lines) of 45 cells (9 spatial frequencies * 5 orientations). **c.** The resulting TvC-functions at a *d'* of 1 (black = most sensitive cell, grey = weighted average).

So far, we have considered a possible way to relate single cell behaviour to psychophysical performance in a contrast discrimination task. More interesting is the question to what degree this model succeeds where others fail. First, and shortly, when it comes to detection in noise, this model predicts a dipper-shaped threshold function (see e.g. Simulation 1). However, this would also be the case for the classical divisive inhibition models if they would be adapted to be able to deal with stochastic stimuli. Second, and more important, this model successfully simulates the effects of noise addition to a contrast discrimination task. Bearing in mind that our weighting rule uses a signal-to-noise ratio statistic, it is clear that the effects of white and spectrally non-flat luminance noise will differ. When white noise is added, each cells signal-to-noise ratio drops in a similar fashion. Consequently, the distribution of weights remains more or less the same. Despite the changes in the weighted average response function (the dynamic range drops, the slope drops and the whole function is shifted to the right), apart from a shift on the axes, not that much changes in the pedestal effect when white noise is added. This is consistent with human data. The story differs for notched noise centred on the signal frequency. In this case, only the signal-to-noise ratio of the off-cells worsens. Consequently, the distribution of weights changes seriously: depending on the characteristics of the notched noise (centre and width of the notch and power of the noise), the decision is almost completely based on the responses of the most sensitive cell. As can be seen in Figure 11c, this leads both to an overall threshold rise and removal of the dip. Again, this is consistent with human data.

Conclusion

Recent work has questioned the validity of the current set of models of human spatial vision. Furthermore, these models are often hard to reconcile with one another and sometimes suffer from an uncomfortable distance to physiological data. Here, the framework of a possible new model which might overcome these problems is presented. The central idea is to combine a Heeger-like front end with a Geisler-like population output decision rule. This results in a model which reconciles constant tuning of the building blocks, as has been postulated often for cortical cells, with dynamic tuning properties of the psychophysical channels. The ultimate goal is to translate this idea in one 2D-input model which aims to explain detection, discrimination, pattern masking and noise masking. The first step towards this goal is formal quantification. While doing so, many issues will have to be studied in

detail (e.g. effects of mild parameter variability in the building blocks, effects of correlated noises, effects of decision rules, effects of stimulus sampling, etc.). It is to be expected that simulations alone will not be sufficient, but that some additional experimental work will be necessary as well, to constrain predictions and avoid inflation of the number of free parameters. The next and crucial step is a massive fit to all our data and to lots of other data sets that are comparable enough (e.g. the data described in Wichmann, 1999). This line of work should at least enable us to judge whether any static contrast perception model can explain human performance in contrast perception studies.

References

Albrecht, D. G., & Devalois, R. L. (1981). Striate cortex responses to periodic patterns with and without the fundamental harmonics. *Journal of Physiology, 319*(1), 497-514.

Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience 7*, 531-546.

Albrecht, D. G., Geisler, W. S., & Crane, A. M. (2003). Nonlinear properties of visual cortex neurons: temporal dynamics, stimulus selectivity, neural performance. In L. Chalupa & J. Werner (Eds.), *The Visual Neurosciences* (pp. 747-764). Boston: MIT Press.

Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research, 27*(11), 1915-1924.

Bird, C., 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*, 1267-1273.

Blackwell, H. R. (1946). Contrast thresholds of the human eye. *Journal of the Optical Society of America 36*(11), 624-643.

Blackwell, K. T. (1998). The effect of white and filtered noise on contrast detection thresholds. *Vision Research 38*(2), 267-280.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology (London) 203*, 237-260.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision 10*, 433-436.

Campbell F. W., Carpenter, R, & Levinson, J. (1969). Visibility of aperiodic patterns compared with that of sinusoidal gratings. *Journal of Physiology (London) 204*, 283-298.

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

Campbell, F. W., & Gubisch, R. W. (1966). Optical quality of the human eye. *Journal of Physiology (London) 186*, 558-578.

Campbell, F. W., & Robson, J. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London) 197*, 551-556.

Carandini, M., Heeger, D., & Movshon, J. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience 17*(21), 8621-8644.

Carandini, M., Heeger, D., & Senn, W. (2002). A synaptic explanation of suppression in visual cortex. *Journal of Neuroscience, 22*(22), 10053-10065.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J, Dan, Y., Olshausen, B. A., Gallant, J. L., & Rust, N. C. (2005). Do we know what the early visual system does? *Journal of Neuroscience 25*(46), 10577-10597.

Carter, B. E., & Henning, G. B. (1971). The detection of gratings in narrow-band visual noise. *Journal of Physiology (London) 219*, 355-365.

Derrington, A. M., & Henning, G. B. (1989). Some observations on the masking effect of two-dimensional stimuli. *Vision Research, 29*(2), 241-246.

DeValois, R. L., & DeValois, K. K. (1988). *Spatial Vision.* Oxford University Press.

Douglas, J. K., Wilkens, L. A., Pantazelou, E. & Moss, F. (1993). Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature 365*, 337-340.

Foley, J. (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model. *Journal of the Optical Society of America A 11*(6), 1710–1719.

Foley, J., & Legge, G. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research 21*, 1041-1053.

Freeman, T. C. B., Durand, S., Kiper, D. C., & Carandini, M. (2002). Suppresion without inhibition in visual cortex. *Neuron 35*, 759-771.

Geisler, W. S., & Albrecht, D. G. (1995). Bayesian analysis of identification performance in monkey visual cortex: Nonlinear mechanisms and stimulus certainty. *Vision Research 35*(19), 2723-2730.

Geisler, W. S., & Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination, and identification. *Visual Neuroscience 14*(5), 897-919.

Gorea, A., & Sagi, D. (2001). Disentangling signal from noise in visual contrast discrimination. *Nature Neuroscience 4*(11), 1146-1150.

Graham, N. (1989). *Visual Pattern Analyzers*. Oxford University Press.

Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: a comparison of single channel and multichannel models. *Vision Research 11*, 251-259.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics.* New York: John Wiley & Sons, Inc.

Heeger, D. (1992a). Normalization of cell responses in cat striate cortex. *Visual Neuroscience 9*, 181-197.

Heeger, D. (1992b). Half-squaring in responses of cat striate cells. *Visual Neuroscience 9*, 427-443.

Henning, G. B. (1988). Spatial-frequency tuning as a function of temporal frequency and stimulus motion. *Journal of the Optical Society of America A 5*, 1362-1373.

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

Henning, G. B., Hertz, B. G., & Hinton, J. (1975). Effects of different hypothetical detection mechanisms on the shape of spatial-frequency filters inferred from masking experiments: I. noise masks. *Journal of the Optical Society of America 71*, 574-581.

Henning, G. B., & Wichmann, F. A. (2007). Some observations on the pedestal effect. *Journal of Vision 7*(1:3), 1-15.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *The Journal of Physiology 160*, 106-154.

Legge, G., & Foley, J. (1980). Contrast masking in human vision. *Journal of the Optical Society of America 70*(12), 1458-1471.

Legge, G. (1981). A power law for contrast discrimination. *Vision Research 21*(4), 457-467.

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A 4*(2), 391-404.

Levin, J. E., & Miller, J. P. (1996). Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. *Nature 380*, 165-168.

Majaj, N. J., Pelli, D. G., Kurshan P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification, *Vision Research 42*, 1165-1184.

Miller, K. D., & Troyer, T. W. (2002). Neural noise can explain expansive, power-law nonlinearities in neural response functions. *Journal of Neurophysiology 87*, 653-659.

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: Discrimination may be better than detection. *Vision Research 14*(10), 1039-1042.

Nagaraja, N. S. (1964). Effect of luminance noise on contrast thresholds. *Jounal of the Optical Society of America 54*, 950-955.

Oruç, I., Landy, M. S., & Pelli, D. G. (2006). Noise masking reveals channels for second-order letters. *Vision Research 46*, 1493-1506.

Patel, A., & Kosko, B. (2005). Stochastic resonance in noisy spiking retinal and sensory neuron models. *Neural Networks 18*, 467-478.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America. A 2*(9), 1508-1532.

Pelli, D. G. (1991). Noise in the visual system may be early. In: M. Landy & J. A. Movshon (Eds.), *Computational Models of Visual Processing* (pp. 147-152). Cambridge: MIT Press.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A 16*, 647-653.

Russell, D. F., Wilkens, L. A., & Moss, F. (1999). Use of behavioural stochastic resonance by paddle fish for feeding. *Nature 402*, 291-294.

Schade, O. H. (1956). Optical and photoelectric analogue of the eye. *Journal of the Optical Society of America 46*, 721-739.

Solomon, J. A. (2000). Channel selection with non-white noise masks. *Journal of the optical society of America A 17*(6), 986-993.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature 369*, 395-397.

Stromeyer, C. F., Klein, S. (1974). Spatial frequency channels in human vision as assymetric (edge) mechanisms. *Vision Research 14*, 1409-1420.

Talgar, C. P., Pelli, D. G., & Carrasco, M. (2004). Covert attention enhances letter identification without affecting channel tuning. *Journal of Vision 4*, 22-31.

Vogels, R., Spileers, W., & Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research 77*(2), 432-436.

Wichmann, F. A. (1999). *Some aspects of modelling human spatial vision: Contrast discrimination*. Unpublished doctoral dissertation, The University of Oxford, Oxford, UK.

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

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics 63*, 1314-1329.

Wickens, T. D. (2002). *Elmentary signal detection theory*. Oxford University Press, Inc.

Wiesenfeld, K., & Moss, F. (1995). Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDS. *Nature 373*, 33-36.

Yang, J., & Makous, J. (1995). Modeling pedestal experiments with amplitude instead of contrast.*Vision Research 35*(14), 1979-1989.

Appendix

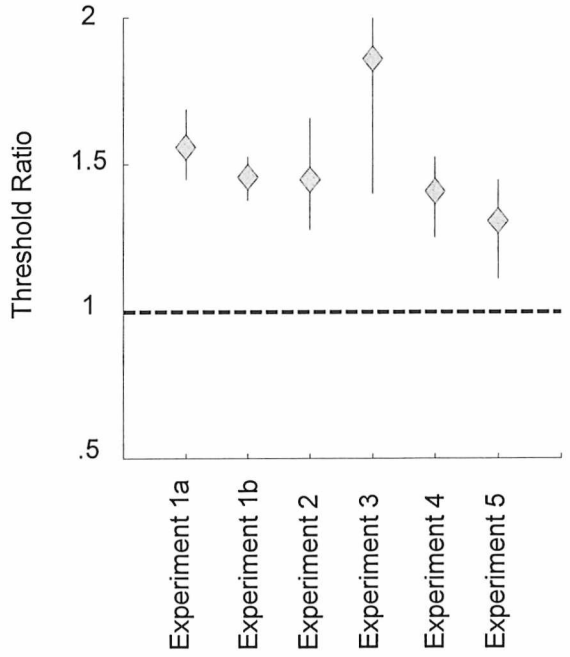
General

Thresholds

All threshold estimates and standard errors were estimated by the procedure described in Wichmann and Hill (2001a,b). Thresholds were estimated at 69% in Experiment 1 and at 75% in Experiment 2. Error bars in the graphs always indicate plus and minus one standard error across subjects. Confidence interval estimates of the improvement ratios were estimated across subjects by means of 10,000 bootstrapped threshold estimates for each subject, as described in Wichmann and Hill (2001a,b). All noise improvement ratios are depicted in Appendix Figure 1. In all the experiments reported here independent noise was generated on each presentation of a stimulus. All experiments were run on Power Macintosh G4 computers, by using the software packages MATLAB (Mathworks, Natick, MA) and PSYCHTOOLBOX (Brainard, 1997). Luminance in all experiments was linearised by means of an 8-bit look-up table.

Definitions

- Contrast: The (Weber) contrast of a pixel is defined as the luminance minus the mean luminance, divided by the mean luminance. The Michelson contrast of an image is defined as the maximal luminance minus the minimal luminance divided by their sum.
- Energy: the squared (Weber) contrast of all pixels in an image, integrated across space, expressed in visual degrees squared.
- Noise power spectral density: the contrast variance of the luminance times the pixel area, expressed in visual degrees squared. It represents the average power at the different frequencies present in the noise.
- *d'*: the difference of two distributions divided by their (common) standard deviation. In psychophysical experiments, the difference of the Z-scores of the hits and false alarms is computed, based on the subjects' responses. As opposed to percentage correct, this measure is insensitive to response bias.



Appendix Figure 1
The maximal noise induced threshold reduction, plus and minus one standard error, for all reported experiments (all thresholds at a *d'* of 1).

Methods Part I

Experiment 1A

Three subjects (I.K., R.G., and B.S.) with normal or corrected-to normal vision participated in this experiment. All except R.G. were naïve concerning the hypothesis tested. Stimuli were presented on a SONY Trinitron GDM-F500R monitor, with a spatial resolution of 1024 * 768 pixels and a temporal resolution of 85 Hz. Mean luminance was 79.5 cd/m^2. Viewing distance was 66 cm, leading to a pixel-size of .033° of visual angle. The stimulus was either a Gabor stimulus (standard deviation of Gaussian envelope: .98°) with a spatial frequency of approximately 2.8 c/deg oriented 25° to the right (signal) from vertical or a blank field (no-signal). Five contrast levels between .3 and .7% Michelson contrast were used. Nine levels of white luminance noise were added to the stimuli, ranging between 0 and 18 * 10^{-7} deg^2 noise power spectral density. Stimulus size was approximately 6.5 * 6.5° of visual angle. Each trial started with a red fixation cross (6.5 × 6.5' of visual angle) for about 176 ms. This was followed by a blank screen for 256 ms. The stimulus was then shown for 118 ms, followed by a blank screen. Probabilities of both stimulus types (target and distracter) were equal (50%) and subjects had to indicate whether they saw the target by pressing '1' and pressing '2' otherwise. Response time was unlimited and no accuracy feedback was given. A block of 100 trials contained only one particular noise and contrast level, assigned randomly to each block. Each subject completed 400 trials per noise and contrast level, leading to 18,000 trials in total per subject.

Experiment 1B

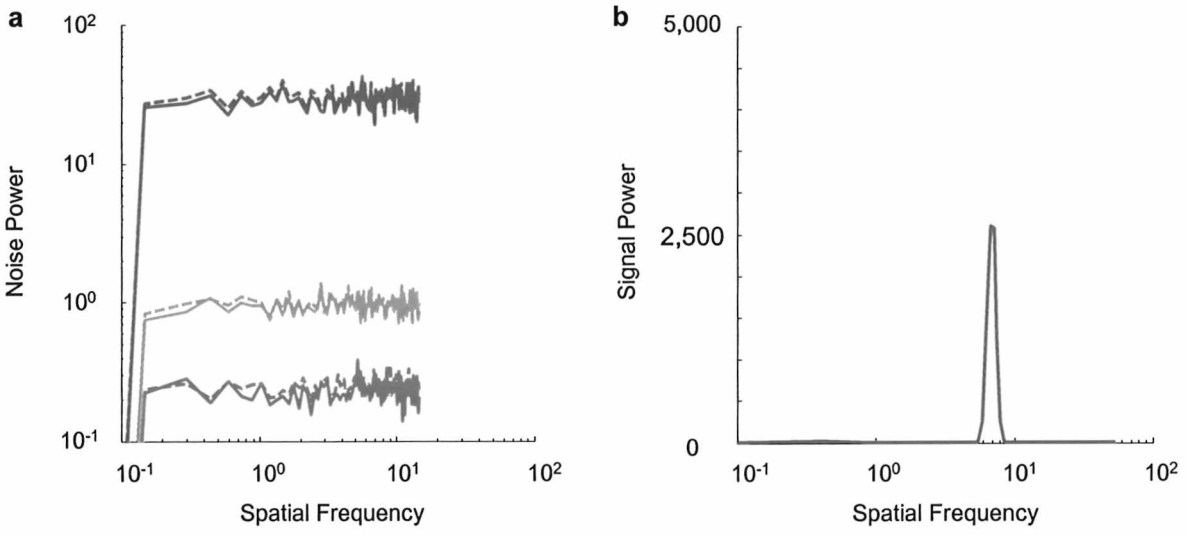
Three subjects (E.G., T.T., and R.G.) with normal or corrected-to normal vision participated in this experiment. All except R.G. were naïve. All parameters were the same as in Experiment 1A, except that the left oriented Gabor stimulus (-45°) was defined as the target stimulus and the right oriented Gabor stimulus (45°) as distracter.

Experiment 2

Three subjects with normal or corrected-to-normal vision (I.P., H.H. and E.G.), all naive, participated in this experiment. Subject E.G. in this experiment was not the same as subject E.G. in Experiment 1B. Stimuli were presented on a SONY Trinitron GDM-FW900 monitor. Screen resolution was 1920 * 1440 pixels and viewing distance 120 cm, leading to a pixel size of .009° of visual angle. The temporal screen resolution was 75 Hz. Mean luminance was 42 cd/m^2. The stimulus was either a Gabor stimulus (standard deviation of Gaussian envelope: .27°) with a spatial frequency of 7 c/deg oriented to the right (45°, signal) or to the left (-45°, no-signal), with a spatial extent of 2.35° of visual angle. Seven contrast levels between 3 and 10% Michelson contrast were used (differing slightly across subjects). Ten levels of white luminance noise were added to the stimuli, ranging between 0 and 38 * 10^{-7} deg^2 noise power spectral density. Stimulus presentation time was approximately 25 ms. Presentation time was doubled for subject E.G. to allow her to reach similar thresholds as the other subjects. On each stimulus presentation, the stimulus was surrounded by a red square (2.35°) at maximal contrast. This square appeared and disappeared with the stimulus. A temporal two-alternative-forced-choice (2AFC) task was used. Each trial started with the presentation of a grey field for about 266 ms. The first stimulus presentation was followed by an ISI of 466 ms, then followed by the second stimulus presentation. Response time was limited to 1,000 ms, and indicated by a grey field of mean luminance, presented after the second stimulus presentation. After the response screen, a screen with a green square (2.35°) appeared for about 750 ms. During this presentation, auditory feedback indicated which interval contained the signal stimulus (low tone of 1,000 Hz: first interval; high tone of 2,000 Hz: second interval). Missed trials (about 2% of the trials) were shown again, but the interval containing the signal stimulus was selected randomly again, and the noise was also refreshed again. Within each block of 50 trials, all variables were randomized. During an entire block a red square of 3.5° of visual angle, surrounding the stimuli, at half of the maximal contrast was presented all the time to ensure eye-accommodation. Subject I.P. and E.G. completed 7,000 trials in total; subject H.H. completed 4,000 trials in total.

Controls

Because of monitor distortions, stimuli may be deformed at low contrasts. Such distortions may, in principle, have caused some of the effects observed, especially in Experiment 1 in which contrasts were very low (in Experiment 2, contrast detection thresholds were around 5%, far above the limits of any video card). To rule out such an artefact explanation, we investigated whether a simulated ideal observer could perform better at noise levels differing from zero because of monitor distortions. The estimated effective images were used as input for an ideal observer that cross-correlated these images with an ideal template of the target. Effective luminance values at the monitor's output were estimated by inserting the theoretical luminance values in the linearization look-up table, rounding them (0-255), and passing them through the monitor's gamma function. Under a wide variety of contrasts and noise levels, performance for such an ideal observer always decreased as noise increased, both for sine- and square-wave stimuli, in contrast to our behavioural findings. In a similar way, we also verified whether the appropriate noise powers were preserved, which was the case (see **Appendix Figure 1**). Finally, at low contrasts, Gabor stimuli may be converted into square wave stimuli, and the noise addition may affect this distortion differentially for different noise levels. These distortions could potentially improve human performance. To rule out such an artefact explanation, we ran a control experiment in which three subjects (two naive) had to detect square-wave stimuli. The experimental set-up was the same as the one in Experiment 1 but the viewing distance was larger (120 cm). Only one high, one low, and a zero noise level were used, but two different fundamental frequencies were used (2.8 and 13.9 c/deg). In this experiment, contrast detection was improved at the low non-zero noise level in all six instances (3 subjects × 2 frequencies, see **Figure 3f**).



Appendix Figure 2

a. The average power spectra of the vertically oriented components of 100 noise images on double logarithmic coordinates. The dotted lines represent theoretical images. The full lines represent estimates of the corresponding effective images. The luminance values of these images were adjusted by the linearization look-up table, rounded to 8 bit (0-255), and passed through the monitor's gamma function. Three noise values are depicted: the lowest noise power used in Experiment 1 (red), the maximal facilitating noise power of Experiment 1 (green), and the highest noise power used in Experiment 1 (blue). **b.** The estimated effective power spectrum of the signal in Experiment 2 at detection threshold on semilogarithmic coordinates. Due to the 8-bit dynamic range, the 1 log unit lower detection thresholds in Experiment 1 result in signals that have a more square-wave like character. Despite these signal differences, results of Experiment 1 and 2 are remarkably similar.

Methods Part II

Experiment 3

Two subjects (A.C. and I.K., both naïve), with normal or corrected-to-normal vision participated in this experiment. Apparatus settings were identical to Experiment 2. The stimulus was either a Gabor stimulus (standard deviation of Gaussian envelope: .27°) with a spatial frequency of 7 c/deg oriented vertically (signal) or a blank field (no-signal). Five contrast levels between 1 and 4% Michelson

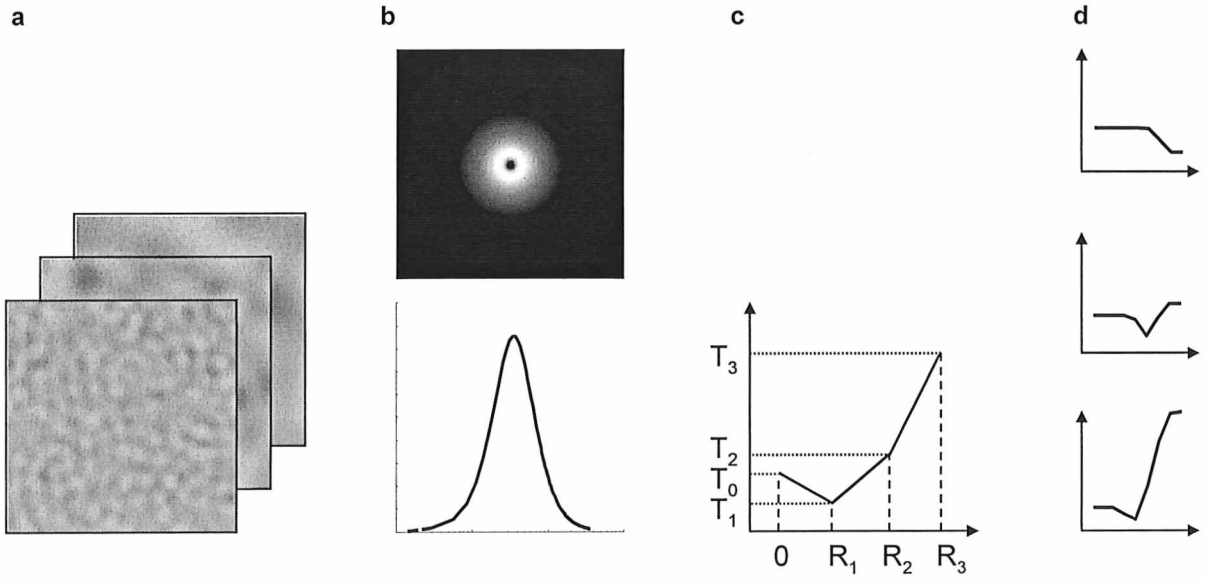
contrast were used. Six levels of low-pass luminance noise with a cut-off frequency of 2.8 c/deg were added to the stimuli, ranging between 0 and 10^{-3} deg^2 noise power spectral density. Stimulus size was approximately 2.35 * 2.35° of visual angle. Stimulus presentation time was approximately 25 ms. A temporal two-alternative-forced-choice (2AFC) task with an interstimulus interval (ISI) of 667 ms was used. The first interval was indicated by an auditory tone of 1,000 Hz, and the second interval by an auditory tone of 2,000 Hz. Each subject completed 100 trials per noise and contrast level, leading to 3,000 trials in total. All variables were randomized within each block of 100 trials.

Experiment 4

Four subjects (A.C., I.P., L.H. and E.G., all naïve), with normal or corrected-to-normal vision participated in this experiment. Apparatus settings were identical to Experiment 2. The stimulus was either a Gabor stimulus (standard deviation of Gaussian envelope: .27°) with a spatial frequency of 5.5 c/deg oriented vertically (signal) or a blank field (no-signal). All other stimulus parameters were identical to Experiment 3. Subjects first completed a pilot experiment in which their threshold vs. white noise power curve was estimated by an adaptive procedure. Based on these results, three noise levels were selected for each observer: one that allowed maximal facilitation, one that would lead to a threshold comparable to their noiseless detection threshold and one that would lead to inhibition. These noise levels differed slightly over participants. In the actual experiment, the noise was low-pass filtered at 8 cut-off frequencies, logarithmically symmetric around the signal frequency (0, 1.57, 2.38, 3.62, 5.5, 8.3, 12.67, and 19.25 c/deg). For each threshold, five contrast levels were used. Task and procedure were identical to Experiment 3. All subjects completed 9,000 trials, except for A.C., who completed 12,000 trials.

Model and Fitting

The procedure described in 'simulation 2' is partly depicted in Appendix Figure 3.



Appendix Figure 3

a. For each fitting procedure, one set of estimates of the effective noise images used in Experiment 4 was generated. To account for the variability inherent to noise images, 25 instances of each condition were used. This leads to a total of 525 images (7 cut-offs * 3 noise levels * 25 images). **b.** The channels in this simulation were circular symmetric filters (**above**: filter gain in Fourier space on a polar plot, **below**: filter gain of the same filter as a function of spatial frequency on semilogarithmic coordinates). Peak sensitivity, bandwidth and slope of both left- and right filterside wer varied as a function of badness of fit in the fitting procedure. Three different types of filters were used: Filters that rectified, squared or double squared the Fourier components before intergrating them over the two spatial dimensions. **c.** For each filter, a nonlinear function relating filter response-to-noise to Energy threshold was specified by interpolating between four data points (four thresholds vs. the average filter responses to the four relevant noise conditions). Because both 0 and the strongest noise condition at the highest cut-off level were included, the whole range of possible filter responses was covered. **d.** In the final stage, thresholds for all other conditions given this particular filter and nonlinear function were estimated.

Methods Part III

Experiment 5

Three subjects with normal or corrected-to-normal vision (L.V., E.G. and L.V.E.), all naive, participated in this experiment. Apparatus settings were identical to Experiment 2. The stimulus was a Gabor stimulus (standard deviation of Gaussian envelope: .27°) with a spatial frequency of 7 c/deg oriented horizontally, with a spatial extent of 2.35° of visual angle. Three levels of white luminance noise, differing slightly over participants, were added to the stimuli: no noise, weak (facilitating) noise and strong (inhibitory) noise. The latter two values were based on a pilot measurement of the Threshold-vs-Noise curve for each participant (2000 trials). For each noise level, 10 pedestal contrast levels between 1 and 40% Michelson contrast were used. Each pedestal contrast was combined with 6 contrast increments. Stimulus presentation time was approximately 50 ms for subjects L.V. and E.G. Presentation time was halved for subject L.V.E. to allow her to reach similar thresholds as the other subjects. On each stimulus presentation, the stimulus was surrounded by a red square (2.35°) at maximal contrast. This square appeared and disappeared with the stimulus. A temporal two-alternative-forced-choice (2AFC) task was used. Each trial started with the presentation of a grey field for about 266 ms. The first stimulus presentation was followed by an ISI of 466 ms, then followed by the second stimulus presentation. Response time was limited to 1,000 ms, and indicated by a grey field of mean luminance, presented after the second stimulus presentation. After the response screen, a screen with a green square (2.35°) appeared for about 750 ms. During this presentation, auditory feedback was given (low tone of 1,000 Hz: incorrect response; high tone of 2,000 Hz: correct response). Missed trials were shown again, but the interval containing the signal stimulus was selected randomly again, and the noise was also refreshed again. Within each block of 50 trials, all variables were randomized. During an entire block a red square of 3.5° of visual angle, surrounding the stimuli, at half of the maximal contrast was presented all the time to ensure eye-accommodation. Subject L.V. and E.G. completed 9,000 trials in total; subject L.V.E. completed 4,500 trials in total.

Model and fitting

The descriptive model fitted to the data of Experiment 5 was introduced by Wichmann (1999) as a generic model of contrast processing because it encompasses three of the main models that have been suggested to account for human contrast discrimination: energy models, nonlinear transducer models and models incorporating divisive contrast-gain control. Observer's responses are modelled within the framework of signal detection theory (SDT). Three hypothetical stages are specified in SDT-models: First, a *stimulus theory* describes how a transduction mechanism maps physical world stimuli to internal states, second, a probabilistic *theory of internal states* describes the probability distribution of the internal states that results from repeated presentation of the same stimulus, and finally a deterministic *response theory* describes a decision rule that maps internal states to a response. In the generic model, the transduction mechanism was operationalised by the generalized four parameter Naka-Rushton function (free parameters α, β, η, and κ). Three additional free parameters were added to describe the internal noise, made up of both a signal-dependent and signal-independent component (free parameters σ, ζ, and γ). It was assumed that the observer's response ('interval 1' or 'interval 2') was the stimulus interval that led to the highest internal state. The model equations were arranged to express percent correct ($f(\Delta x, x)$) as a function of the contrast increment (Δx) and the pedestal contrast (x).

$$f(\Delta x, x) = \int_0^\infty \left[\sqrt{2\pi \operatorname{var}(\Delta x, x)}\right]^{-1} e^{-\frac{(z-mean(\Delta x,x))^2}{2\operatorname{var}(\Delta x,x)}} dz,$$

where z is a dummy variable, and $mean(\Delta x, x)$ and $var(\Delta x, x)$ are given by

$$mean(\Delta x, x) = \alpha\left(\frac{(\Delta x + x)^\eta}{\beta + (\Delta x + x)^\kappa} - \frac{x^n}{\beta + x^\kappa}\right),$$

and

$$\mathrm{var}(\Delta x, x) = 2\sigma^2 + \zeta^2\left(\left(\frac{\alpha(\Delta x + x)^{\eta}}{\beta + (\Delta x + x)^{\kappa}}\right)^{2\gamma} + \left(\frac{\alpha x^{\eta}}{\beta + x^{\kappa}}\right)^{2\gamma}\right).$$

One of the seven parameters can be arbitrarily set to any value. To adopt convention, σ was taken to be 1. An additional (highly constrained) free parameter λ ("lapse rate") was introduced to avoid biased parameter estimates. To find the surface $f(\Delta x, x)$ that maximizes the likelihood that the dataset were generated from a process with success probability given by $f(\Delta x, x)$, the log-likelihood of the surface $f(\Delta x, x)$ given the parameters (α, β, η, κ, σ, ζ, γ and λ) was maximized using purpose-written software in MATLAB.