# Modelling contrast discrimination data suggest both the pedestal effect and stochastic resonance to be caused by the same mechanism

**Robbe L. T. Goris**

Laboratory for Experimental Psychology, University of Leuven, Tiensestraat, Leuven, Belgium

**Johan Wagemans**

Laboratory for Experimental Psychology, University of Leuven, Tiensestraat, Leuven, Belgium

**Felix A. Wichmann**

Technische Universität Berlin, FG Modellierung Kognitiver Prozesse, Sekr. FR, Franklinstr., Berlin, FRG and Bernstein Centre for Computational Neuroscience, Philippstr. Haus, Berlin, FRG

Computational models of spatial vision typically make use of a (rectified) linear filter, a nonlinearity and dominant late noise to account for human contrast discrimination data. Linear–nonlinear cascade models predict an improvement in observers' contrast detection performance when low, subthreshold levels of external noise are added (i.e., stochastic resonance). Here, we address the issue whether a single contrast gain-control model of early spatial vision can account for both the pedestal effect, i.e., the improved detectability of a grating in the presence of a low-contrast masking grating, and stochastic resonance. We measured contrast discrimination performance without noise and in both weak and moderate levels of noise. Making use of a full quantitative description of our data with few parameters combined with comprehensive model selection assessments, we show the pedestal effect to be more reduced in the presence of weak noise than in moderate noise. This reduction rules out independent, additive sources of performance improvement and, together with a simulation study, supports the parsimonious explanation that a single mechanism underlies the pedestal effect and stochastic resonance in contrast perception.

Keywords: spatial vision, contrast discrimination, dipper function, stochastic resonance, contrast gain-control model, masking

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

## Introduction

Threshold detection studies have provided much evidence consistent with a model of our visual system in which visual information is analyzed in relatively independent and approximately linear spatial-frequency and orientation-selective filters (Blakemore & Campbell, 1969; Campbell, Carpenter, & Levinson, 1969; Campbell & Robson, 1968; DeValois & DeValois, 1988; Graham & Nachmias, 1971). The corresponding spatial weighting functions of such band-limited channels are slightly asymmetric on double logarithmic coordinates (Henning, 1988; Henning, Hertz, & Hinton, 1981) but are nonetheless typically approximated by Gabor functions. Though not an "optimal" stimulus (Watson, Barlow, & Robson, 1983; but see Henning, Derrington, & Madden, 1983), Gabors are popular stimuli to explore contrast detection because they are band limited in spatial frequency and localized in space (Daugman, 1985). Studies that made use of more complex stimuli than single component gratings have reported results that are difficult to reconcile with the aforementioned independence and linearity (Derrington & Henning, 1989; Henning, Bird, & Wichmann, 2002; Henning, Hertz, & Broadbent, 1975), but the multi-channel model still captures much of what we understand of early spatial vision.

To gain insight in visual processing at supra-threshold contrasts—a prerequisite for any model of spatial vision to be truly useful—sinusoidal contrast discrimination has been studied extensively (e.g., Baker, Meese, & Georgeson, 2007; Bird, Henning, & Wichmann, 2002; Foley, 1994;

doi: 10.1167/8.15.17 Received January 28, 2008; published November 24, 2008 ISSN 1534-7362 

Foley & Chen, 1997; Foley & Legge, 1981; Georgeson & Georgeson, 1987; Gorea & Sagi, 2001; Henning & Wichmann, 2007; Kontsevich, Chen, & Tyler, 2002; Legge, 1981; Legge & Foley, 1980; Legge, Kersten, & Burgess, 1987; Nachmias & Sansbury, 1974; Wichmann, 1999; Yang & Makous, 1995). The main finding of these studies is the so-called pedestal effect, i.e., the improved detectability of a grating in the presence of a low-contrast "pedestal" stimulus of identical spatial frequency, phase, and orientation. At higher pedestal contrasts, discrimination thresholds rise in a way that is roughly consistent with Weber's law. Analogs of the pedestal effect in spatial vision have been reported in other sensory modalities such as amplitude discrimination in hearing and, recently, flicker discrimination in vision (Smithson, Henning, MacLeod, & Stockman, in press).

The mechanisms suggested to underlie the dipper-shaped thresholds-vs.-contrast (TvC) function in spatial vision include nonlinear transduction (e.g., Legge & Foley, 1980), contrast gain-control (e.g., Foley, 1994; Foley & Chen, 1997), and uncertainty reduction (e.g., Pelli, 1985). While the first two types of models—indistinguishable for single component sinusoidal contrast discrimination (Wichmann, 1999)—effectively make use of an expansive nonlinearity in a sensory post-filter stage to explain the pedestal effect, some uncertainty models leave the sensory processing linear but place the nonlinearity, required to explain the dipper function, in the decision process. In effect, all models that explain the dipper function with some degree of success make use of a (rectified) linear filter stage, a nonlinear post-filter stage, and a dominant late internal noise source.

Recently, Goris, Zaenen, and Wagemans (2008) noted that such linear–nonlinear cascade models predict an improvement in observers' performance when low, subthreshold levels of external noise are added. The improvement arises because rectification causes the external noise to increase the mean filter response to both a weak signal-plus-noise and a noise stimulus. Because these responses are expanded subsequently, the difference between the means of the internal response distributions representing the noise and signal-plus-noise increases. If the crucial performance-limiting noise source is located in later processing stages, this increased difference between the means will lead to a better signal-to-noise ratio at the decision stage, resulting in improved contrast detection.

Consistent with this prediction, Goris et al. (2008) showed that detection thresholds reach a minimum for very low noise levels. This type of phenomenon—improved signal transmission due to the addition of external noise—is termed stochastic resonance in physics (Wiesenfeld & Moss, 1995). Similar observations have been reported once before in spatial vision (Blackwell, 1998) and in other sensory modalities such as tactile (Collins, Imhoff, & Grigg, 1996) and auditory perception (Zeng, Fu, & Morse, 2000). To test their speculation that the mechanism underlying the pedestal effect also leads to stochastic resonance, Goris et al. fitted a contrast gain-control model to their detection-in-noise data and found the fits to be reasonable.

Aiming to explain the pedestal effect, the Goris et al. (2008) gain-control model assumes that the effect characterizes the operation of individual, spatial-frequency tuned channels. Recently, it has been suggested that this is not the case in spatial vision (Henning & Wichmann, 2007). The disappearance of the pedestal effect in notched noise led Henning and Wichmann (2007) to argue that the pedestal effect may be characteristic of the way in which observers use information from the spatial-frequency and orientation-selective channels tuned to frequencies and orientations away from the signal rather than characteristic of a single channel tuned to the spatial characteristics of the signal to be detected. The gain-control model may thus not be an *explanation* of single channel behavior despite offering an excellent *description* of contrast discrimination performance (e.g., Wichmann, 1999). In this paper, we exploit the gain-control model as a powerful statistical tool with few free parameters to describe our contrast discrimination data, and we use the fits to different data sets to make statistically sound inferences about changes in the data.

One way of observing effects of experimental manipulations on contrast discrimination is to normalize all pedestal and signal contrasts by the detection threshold measured in the absence of a pedestal. When this is done, data are expressed on the same scale, irrespective of differences in detection threshold. Once corrected for absolute visibility, the dipper-shaped TvC function has been reported to be remarkably invariant to parameter manipulations such as retinal illumination (e.g., Yang & Makous, 1995), retinal eccentricity (e.g., Bradley & Ohzawa, 1986), varying degrees of stimulus uncertainty (e.g., Foley & Schwartz, 1998), and cross-surround facilitation (Yu, Klein, & Levi, 2002). Measurements in the presence of broadband noise have revealed a similar invariance (e.g., Pelli, 1985), though the depth of the dip is reduced for at least some observers (e.g., Figure 7 in Henning & Wichmann, 2007). This might be a consequence of the additional stimulus variability that is introduced: In the presence of noise, the pedestal stimulus might have an *average* optimal contrast, but on each trial, due to the noise, the contrast in the relevant channel will be somewhat higher or lower and thus not optimal. This will lead to a higher threshold, i.e., a reduced pedestal effect. Considered this way, the invariance to the presence of noise displayed by some observers, especially in 1-D noise (e.g., Figure 8 and 10 in Henning & Wichmann, 2007), may indicate a change in the underlying mechanism. In 2-D noise, the number of active channels, sensitive to the signal, is similarly important: Having few active channels, perhaps even with correlated noise (Henning et al., 2002) could be expected to reduce the size of the pedestal effect due to the additionally introduced stimulus variability. On the other hand, having many active channels that sample different image regions and hence different regions of noise could enable the averaging

out of noise-induced effects of signal variability and thus leave the size of the pedestal effect (almost) unchanged. Whatever the exact mechanism underlying contrast discrimination in noise may be, our main interest here concerns the dipper-effect being approximately invariant to addition of (strong) broadband noise.

In sum, all psychophysical spatial vision models that aim to explain the pedestal effect make use of a (rectified) linear filter stage, a nonlinear post-filter stage, and a dominant late internal noise source. This implies that these models actually predict that stochastic resonance occurs in contrast detection. This prediction has been confirmed experimentally (Goris et al., 2008). Here, we address the issue whether a single model of early spatial vision, having a (rectified) linear filter, a nonlinearity, and dominant late noise can account for both the pedestal effect and stochastic resonance. If the same mechanism underlies the pedestal effect and stochastic resonance, we should expect the pedestal effect to be reduced in the presence of weak noise. The logic of our argument is as follows: The mean response of a linear filter to a low contrast signal is raised in the presence of external noise due to rectification. As explained above, for certain external noise levels, the subsequent nonlinear stage leads to an improved discriminability of the signal in the zero-pedestal contrast case. However, as we shall see later, as pedestal contrast increases from zero, the response-enhancing effect of rectification diminishes and soon disappears completely. From these pedestal contrasts on, the only effect of weak noise is that it introduces additional variability. Thus, at an "optimal" noiseless pedestal contrast, addition of external noise will not lead to a benefit but rather hurt contrast discrimination performance. Therefore, if one mechanism underlies both facilitation phenomena, addition of weak noise can only improve discrimination performance at pedestal contrasts lower than the "optimal" pedestal contrast, implying that the size of the pedestal effect is reduced in the presence of weak noise.

An alternative hypothesis may be to argue that such a reduced dipper effect in weak noise is a consequence of the stimulus variability introduced by the noise (see above). In this view, strong noise (which produces even more stimulus variability) should lead to a stronger reduction in the dipper effect than weak noise does. As our data will bear out, this is not the case however: The reduced dip in weak noise is not simply a consequence of increased stimulus variability.

# Methods

## Equipment

The experiments were run on a Power Macintosh G4 computer using the software packages MATLAB (MathWorks, Natick, MA) and PSYCHTOOLBOX (Brainard, 1997; Pelli, 1997a). Gamma correction using an 8-bit lookup table ensured that the monitor was linear over the entire luminance range used in the experiments. The stimuli were presented on a SONY Trinitron GDM-FW900 monitor with a spatial resolution of 1920 × 1440 pixels and a temporal resolution of 75 Hz. The experiment was run in a darkened room and the screen's mean gray background luminance was 42 cd/m$^2$. Viewing distance was 120 cm, leading to a pixel-size of 0.009° of visual angle.

## Observers

Four observers participated (L.V., E.G., B.B., and L.V.E.). All were well practiced in the task and familiar with the stimuli before data collection began and had normal or corrected-to-normal vision. All observers were naïve to the purpose of the experiment.

## Stimuli

The Gabor stimuli consisted of horizontally orientated sine-gratings with a spatial frequency of 7 c/deg, which were multiplied by a circularly-symmetric two-dimensional spatial Gaussian envelope with a $\sigma$ of 0.27°. Stimuli had a spatial extent of 2.35° of visual angle.

For each stimulus presentation in the noise conditions, a fresh noise sample was generated at every pixel, sampled from independent identically distributed Gaussian distributions centered at the mean luminance. For each observer—in addition to the no-noise condition—two noise conditions were run; for the noises noise-power spectral densities ranging between 0.84 and 42 × $10^{-7}$ deg$^2$ were used. Noise-power spectral density is defined as the luminance variance multiplied by the pixel area, expressed in visual degrees squared. Effects of temporal waveform, duration and bandwidth are not considered here. It is proportional to the average power at the different frequencies present in the noise. The maximal amount of clipping (i.e., pixels set to the minimal or maximal luminance values because of the limited 8-bit dynamic range of the DACs on the video card) at the highest noise level was around 2.5%. Through simulations in MATLAB, this level of clipping was calculated to have no significant influence on the spectral properties of the Gaussian white noise.

## Procedure

The same procedure as described in Goris et al. (2008) was adopted. A temporal two-alternative-forced-choice (2AFC) task was used. Stimulus presentation time was

approximately 12.5 ms for subject B.B., 25 ms for subject L.V.E., and 50 ms for subject E.G. and L.V. We used short presentation times to obtain detection thresholds around 5% Michelson contrast for each observer—due to their different absolute sensitivities, we thus had to use different presentation times for our observers. Signal presentation time has a significant influence on the parameters of the linear–nonlinear model typically used to fit the TvC functions (Wichmann, 1999). However, as all our analyses will be performed on the data of individual observers, differences resulting from the different presentation times are immaterial with respect to the conclusions drawn. We attempted to keep detection thresholds comparatively high at 5% contrast to reduce the potential influence of artefacts due to the nonlinear monitor operations (i.e., power saving features, gamma correction, luminance rounding etc.; Bach, 1997; Bach, Meigen, & Strasburger, 1997; Brainard, 1989; Naiman & Makous, 1992; Pelli, 1997b; Wichmann, 1999). Estimates of the effective images ensured that these nonlinear monitor operations did not significantly affect the spectral properties of the noise stimuli. To minimize potential effects of spatial and temporal stimulus uncertainty, the stimulus was surrounded by a red square (2.35°) at full contrast; this square appeared and disappeared with the stimulus onset and offset.

All trials started with the 266-ms presentation of a gray field of mean luminance. The first stimulus presentation was followed by an inter-stimulus interval (ISI) of 466 ms and then followed by the second stimulus presentation. The ISI was thus about 20 times longer than the stimulus duration and more than three times as long as current estimates of the temporal impulse response of human spatial-frequency channels (Graham, 1989). The noise was turned on and off with both stimulus presentations. Independent noise samples were presented in the two observation intervals. Response time was limited to 1,000 ms and was indicated by a gray 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. Observers received auditory feedback after each trial; missed trials (about 1%) were shown again; on these occasions the interval containing the signal stimulus was selected randomly again and the noise was drawn anew. After a few training blocks, participants were familiar with this steady temporal task profile.

Before data collection began, all observers participated in a detection-in-noise task that consisted of 2,000 trials. These training data were used to estimate the noise level that would evoke maximal facilitation—stochastic resonance—as a function of noise power density for each observer (i.e., the noise level used in the weak-noise condition, which varied somewhat across observers). Across observers, the weak noise had an average power spectral density of $3.9 \times 10^{-7}$ deg$^2$, and the moderate noise of $42 \times 10^{-7}$ deg$^2$, approximately a log unit increase. We call the highest noise level used in this experiment "moderate" because it leads, on average, only to a rise by a factor of 1.34 in the detection threshold at 75% correct. The high detection thresholds purposely enforced through the use of very short presentation times, together with the limited dynamic range of today's graphics cards and CRTs that prevented us from using higher noise power densities, unfortunately prevented us from having stronger masking effects.

Ten pedestal contrast levels, ranging between 0% and 40% were used. To obtain psychometric functions, six increment contrasts were tested at each noise and pedestal level. Within each block of 50 trials, 50 conditions were randomly sampled from all 180 conditions (3 noise power densities × 10 pedestal contrasts × 6 increment contrasts). To help observers to maintain constant eye accommodation, a red square of 3.5° of visual angle surrounding the stimuli was presented throughout the experiment. Subjects L.V. and B.B. completed 9,000 trials in total. Because one data point heavily influenced the initial estimate of subject E.G.'s noiseless detection threshold, she completed 700 additional detection trials and thus 9,700 trials in total. Subject L.V.E. completed only 4,500 trials in total—25 instead of 50 trials per condition—and no ethically acceptable means could be found to motivate her to finish the experiment.

## Results

A summary of the results for all observers is shown in Figure 1. Figure 1a displays the data of all observers in the three detection conditions, i.e., detection without noise and detection in the presence of weak and moderate noise. As for all figures in this paper, red symbols and lines refer to the no-noise condition, green to the weak-noise condition, and blue to the moderate-noise condition. Different symbols refer to different observers. To express the data of all observers on the same scale, the contrasts of the stimuli of each noise condition were first normalized by the 75% correct detection threshold of each observer in that condition and then multiplied by the average detection threshold of all observers in that condition. The psychometric functions relating the percentage of correct responses to the logarithm of signal contrast were fit with Weibull functions using the maximum-likelihood procedure of Wichmann and Hill (2001a, 2001b); 100,000 bootstraps were run to estimate confidence intervals.

As can be seen in Figure 1a, we find stochastic resonance as in Goris et al. (2008): Addition of weak noise improves contrast detection performance. This is not the case for moderate noise, which increased the 75% correct threshold by a factor of 1.34. Psychometric functions measured with and without weak noise added

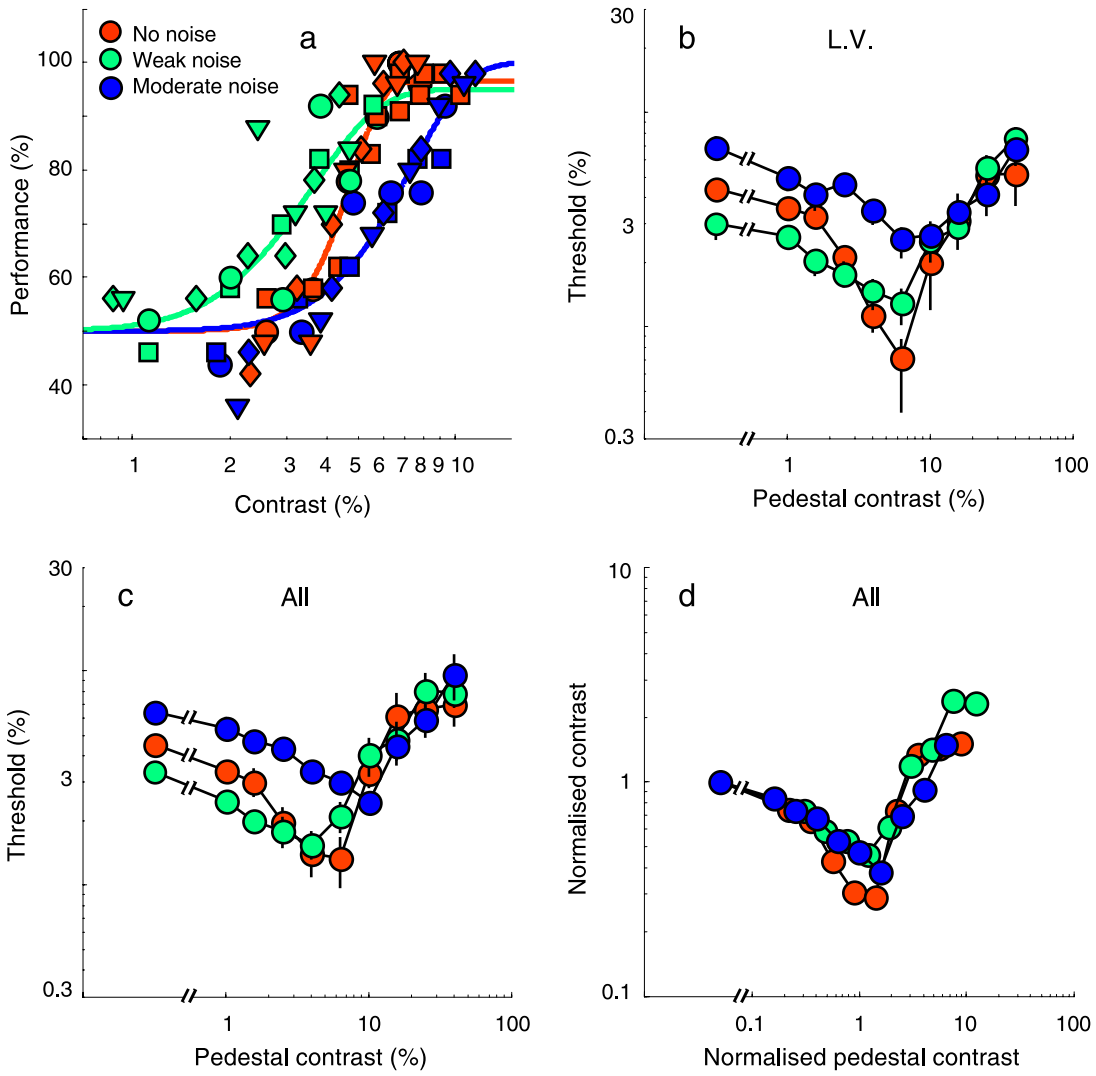


Figure 1. (a) Performance as a function of signal contrast on semi-logarithmic coordinates for the three detection conditions. Red symbols refer to the no-noise condition, green to the weak-noise condition, and blue to the moderate-noise condition. Full lines indicate the best fitting Weibull functions to the pooled data of all observers (L.V.: ●, E.G.: ■, B.B.: ♦, and L.V.E.: ▼). (b) The complete TvC functions at 75% correct for observer L.V. (c) The complete TvC functions at 75% correct averaged over observers. (d) The same functions as in panel c, normalized by their detection thresholds.

to the stimuli are not parallel on semi-logarithmic coordinates. Even though the weak noise is barely visible, the slope of the psychometric function is shallower than in the noiseless detection condition. For the moderate-noise condition, the estimated slope is in-between (and does not differ significantly from the two other conditions given the amount of data collected). This change in slope as a function of noise level is consistent with earlier observations on the slope of psychometric functions in contrast discrimination: Psychometric functions are steepest without any pedestal added and most shallow in the trough of the dip (Nachmias & Sansbury, 1974; Wichmann, 1999). As in contrast discrimination, this implies that the strength of the noise facilitation effect depends on the performance level considered, i.e., the size of the threshold reduction is inversely related to performance level. For our current data, the facilitation at 60% correct is estimated to be of a factor 1.74 (i.e., the ratio of the 60% correct thresholds measured without noise and with weak noise added equals 1.74), at 75% correct this factor is 1.39, and at 90% it is 1.15. Though the particular facilitation factors did vary somewhat over observers (ranging between 1.26 and 1.47 at 75% correct), all individual data sets displayed the trends described above.

We now consider the results at all pedestal contrasts. Figure 1c shows the full TvC functions at 75% correct for all noise conditions. The most leftward points denote detection (i.e., the pedestal contrast equals zero). For illustration purposes, thresholds were simply averaged over observers, without any attempt to rescale the data. The normalized versions of these curves can be seen in Figure 1d, in which threshold and pedestal contrasts were

normalized by the average test contrast threshold in each of the three noise conditions (when the pedestal was zero). Despite the simple averaging of the data, the TvC functions capture the main trends that can be seen in all individual data sets (one example is shown in Figure 1b). Figures 1b and 1c illustrate that at the lowest pedestal levels, the data shown in green lie below the no-noise data, shown in red. For low pedestal contrasts, we thus find stochastic resonance, i.e., contrast discrimination performance benefits from the presence of weak noise. At slightly higher pedestal contrasts, this effect disappears and eventually, around the trough of the dip for the no-noise condition, weak noise hinders performance. We thus find no "super-dipper" in the presence of weak noise, as for instance has been reported for contrast discrimination in combination with cross-surround facilitation (Yu et al., 2002). Stochastic resonance and the pedestal effect are not simply additive and thus not two independent effects, in line with our hypothesis that a single mechanism accounts for both. Furthermore, it can be seen in the normalized plots in Figure 1d that weak noise seems to reduce the pedestal effect to a similar, perhaps even stronger degree as moderate noise does. For these averaged data, maximal threshold reduction at 75% correct was of a factor 3.44 without noise (ranging between 3.15 and 6.13 over observers), 2.17 in the presence of weak noise (ranging between 1.96 and 3.53), and 2.62 in the presence of moderate noise (ranging between 2.28 and 2.99). A stronger reduction of the dipper effect in weak noise than in strong noise is not consistent with the hypothesis that this reduction is solely due to the introduced stimulus stochasticity. However, for all observers, the 95% confidence intervals of the maximal threshold reduction factors of all conditions overlapped, pointing to the need to use more sophisticated statistical techniques to analyze our data (see the modelling below).

In sum, we find stochastic resonance, i.e., a significantly reduced detection threshold in the presence of weak noise. Furthermore, when data of all observers are pooled, the slope of the psychometric function is significantly decreased in the presence of weak noise. The full TvC functions at 75% correct suggest that the pedestal effect is diminished in the presence of both weak and moderate noise, though this reduction is not significant if assessed for each observer independently. The failure to find a "super-dipper" at 75% correct suggests that stochastic resonance and the pedestal effect are *not* independent effects. Of course, only considering 75% correct-thresholds ignores much of the information present in our data set. To study whether and how the presence of noise changes contrast discrimination performance in more detail, we will make use of the standard gain-control model introduced by Foley (1994) and elaborated by Wichmann (1999). In this application, we mainly use this model as a statistical tool to get a full quantitative description of our data with as few parameters as possible.

## Model: Equations, fitting, evaluation and selection

### Equations

We used the standard divisive gain-control model as formalized by Wichmann (1999). Observer's responses are modelled within the framework of signal detection theory (SDT; Green & Swets, 1966). Three hypothetical stages are specified in SDT-models: First, a *stimulus theory* describes how a transduction mechanism maps physical 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 gain-control model, the transduction mechanism was chosen to be the generalized four parameter Naka–Rushton function (free parameters $\alpha$, $\beta$, $\eta$, and $\kappa$). These parameters express the response gain ($\alpha$), the semisaturation contrast ($\beta$), the response exponent ($\eta$), and the gain-control exponent ($\kappa$) of the contrast response function. One additional free parameter was added to describe the internal noise, assumed to be Gaussian and signal independent (free parameter $\sigma$). Recently, there has been some debate in spatial vision regarding the question whether a signal-dependent source might also contribute to internal noise (e.g., Georgeson & Meese, 2006; Gorea & Sagi, 2001; Kontsevich et al., 2002; Wichmann, 1999). Level-dependent noise may be needed to explain contrast discrimination performance at high pedestal levels. Here it was not needed because the majority of data points were gathered at relatively low pedestal levels. Thus, we did not include level-dependent noise for our modelling to reduce the number of free parameters. This is not to argue that level-dependent noise may not be crucial; indeed, one of us showed that level-dependent noise is critically needed to fit contrast discrimination data at high pedestal contrasts (Wichmann, 1999).

As is standard, it was assumed that the observer's response ("interval 1" or "interval 2") is determined by the stimulus interval that led to the highest internal state. The model equations (see Equations 1–3) were arranged to express percent correct, $p(\Delta x, x)$, as a function of the contrast increment ($\Delta x$) and the pedestal contrast ($x$) in a 2AFC-task.

$$p(\Delta x, x) = \int_0^\infty \frac{1}{\sqrt{2\pi g(\Delta x, x)}} e^{-\frac{(z-f(\Delta x,x))^2}{2g(\Delta x,x)}} dz, \quad (1)$$

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

$$f(\Delta x, x) = \alpha\left(\frac{(\Delta x + x)^\eta}{\beta^\kappa + (\Delta x + x)^\kappa} - \frac{x^\eta}{\beta^\kappa + x^\kappa}\right), \quad (2)$$

and

$$g(\Delta x, x) = 2\sigma^2. \quad (3)$$

### Fitting

One of the five parameters can be arbitrarily set to any value. To follow the usual convention, $\sigma$ was taken to be 1, resulting in a four free parameter model. An additional (highly constrained) vector of free parameters $\lambda$ ("lapse rates"), estimated for each noise and pedestal combination as explained by Wichmann and Hill (2001a, 2001b), was introduced in the fitting of the model to avoid biased parameter estimates (for details see Wichmann, 1999). Priors were introduced for each parameter to constrain estimates to realistic values. To find the surface $p(\Delta x, x)$ that maximizes the likelihood that the data were generated from a process with success probability given by $p(\Delta x, x)$, the log-likelihood of the surface $p(\Delta x, x)$ given the parameters ($\alpha$, $\beta$, $\eta$, and $\kappa$) was maximized using purpose-written software in MATLAB (*fminsearch*, which makes use of the Nelder–Mead simplex search method). The log-likelihood of the surface $p(\Delta x, x)$ given parameter vector $\boldsymbol{\theta}$, containing $\{\alpha, \beta, \eta, \kappa\}$ with $\sigma = 1$ and $\boldsymbol{\lambda}$ equal to the lapse rate vector derived from the psychometric function fits, is given by Equation 4:

$$l(\theta) = \sum_{j=1}^{Z}\sum_{i=1}^{K_j} \log\binom{n_{ji}}{y_{ji}n_{ji}} + y_{ji}n_{ji}\,\log\left(p(\Delta x_{ji}, x_j; \theta)\right) + \left(1 - y_{ji}\right)n_{ji}\,\log\left(1 - p\left(\Delta x_{ji}, x_j; \theta\right)\right), \quad (4)$$

with $n_{ji}$ the number of trials (block size) measured at pedestal contrast $j$ and signal contrast $i$ and $y_{ji}$ the proportion of correct responses in that condition. Because the problem is nonconvex due to $\boldsymbol{\lambda}$, a multi-start procedure with semi-randomly chosen initial parameter values was used. For each model fit reported, at least 20 different starting points were used.

### Evaluation

To evaluate model fits, we considered the overall distance between model prediction and data and the presence of systematic errors in the residuals. Quality of the overall fit can be assessed by judging total deviance (see Equation 5), i.e., the log-likelihood ratio of the saturated model and the best fitting model (the saturated model is the model with no residual error between model predictions and data). What deviance does not assess, however, are systematic trends in the deviance residuals (see Equation 6), i.e., the agreement between individual data points and the corresponding model prediction. For binomial data, deviance is expressed by Equation 5,

$$D = 2\sum_{j=1}^{Z}\sum_{i=1}^{K_j}\left\{n_{ji}y_{ji}\,\log\left(\frac{y_{ji}}{p\left(\Delta x_{ji}, x_j\right)}\right) + n_{ji}\left(1-y_{ji}\right)\,\log\left(\frac{1 - y_{ji}}{1 - p\left(\Delta x_{ji}, x_j\right)}\right)\right\}. \quad (5)$$

This statistic indicates how well a model describes data. Asymptotically, it can be shown to be $\chi^2$-distributed, with degrees of freedom equal to the number of data blocks minus the number of free parameters if the model is correct and an observer behaves perfectly stationary during the whole experiment (and thus generates truly binomially distributed data). Often, due to a variety of reasons, this is not the case. Responses of nonstationary observers are more variable than binomially distributed data and thus lead to higher deviances (overdispersion). Wichmann (1999) has shown that, due to the typically relatively small number of measurements, the asymptotically derived deviance distributions often fail to approximate the real deviance distribution for psychophysical data sets. The real deviance distribution can be estimated easily by means of Monte Carlo simulations. As suggested by Wichmann (1999), we estimated the deviance distribution for each model fit by means of 10,000 simulated data sets for an observer whose correct responses in our experiment are binomially distributed as specified by the model fit. From these simulations, we derived critical values for each reported fit. These values indeed often deviate in an unpredictable manner from the asymptotically derived values, confirming Wichmann (1999). Of course, these critical values do not take into account the nonstationariness of real observers. Overdispersion may thus still occur.

Each deviance residual $d_{ji}$ is defined as the square root of the deviance value calculated for data point $i$ in isolation, signed according to the direction of the arithmetic residual $y_{ji} - p(\Delta x_{ji}, x_j)$. For binomial data, this is expressed by Equation 6,

$$d_{ji} = \mathrm{sgn}\left(y_{ji} - p\left(\Delta x_{ji}, x_j\right)\right) \times \sqrt{2\left[n_{ji}y_{ji}\log\left(\frac{y_{ji}}{p\left(\Delta x_{ji}, x_j\right)}\right) + n_{ji}\left(1-y_{ji}\right)\log\left(\frac{1-y_{ji}}{1-p\left(\Delta x_{ji}, x_j\right)}\right)\right]}. \quad (6)$$

Note that $D = \sum_{j=1}^{Z}\sum_{i=1}^{Kj} d_{ji}^2$, as for RMSE. Systematic trends in deviance residuals indicate a systematic misfit of the model.

### Selection

Model selection refers to the problem of selecting, from a group of competing models, the model that best predicts future data, i.e., that generalizes best. Due to noisiness of the data and the problem of over-fitting, this is not simply a matter of goodness-of-fit. In order to select the model with highest *predictive accuracy*, different quantitative methods have been suggested (for a review and overview, see Myung, 2000; Pitt & Myung, 2002; Pitt, Myung, & Zhang, 2002; Wasserman, 2000; Zucchini, 2000). As there is no generally agreed consensus as to what method is best, we used three different model selection criteria: *Akaike*'s *information criterion* (AIC), *Bayesian information criterion* (BIC), and *cross-validation* (CV). AIC trades simplicity and goodness-of-fit for nested models. It is commonly formulated for model family $\mathbf{F}$ as given by Equation 7.

$$\mathrm{AIC}(\mathbf{F}) = D + 2l, \tag{7}$$

With $l$ the number of adjustable parameters.[1] As for all model selection methods mentioned, the model that minimizes the criterion should be selected. AIC has the additional advantage that, for nested models, the reduction in AIC can be compared to a $\chi^2$-distribution with degrees of freedom equal to the difference in number of free parameters between the models.

For BIC the complexity measure is not only sensitive to the number of adjustable parameters, but it is also modified by sample size (see Equation 8). Thus, BIC is stricter than AIC for $\ln(n) > 2$, i.e., $n > 7$.

$$\mathrm{BIC}(\mathbf{F}) = D + l\ln(n). \tag{8}$$

In machine learning, the standard model selection criterion is CV (Hastie, Tibshirani, & Friedman, 2003; for its use in psychology, see e.g., Browne, 2000). We used 10-fold CV: Here the data are divided into ten subsamples of equal size. The model is fitted to nine subsamples, the training set. The normalized deviance of the model fit to the training set is called *training error*. The normalized deviance of the same parameter estimates to the subsample that was left out during parameter estimation (i.e., the test set) is called *test error* (see Equation 9). By minimizing test error, CV has a strong and intuitively appealing emphasis on generalizability (and large differences between training and test error are indicative of over-fitting).

$$\mathrm{CV}(\mathbf{F}) = D_{\mathrm{Test}}. \tag{9}$$

In the fitting of each training set, lapse rate $\lambda$ was re-estimated for each pedestal and noise combination. Each subsample can be used once as test set, which results in ten estimates of test error. Assuming stationary data, i.e., training and test data come from the same distribution, a model that is correct—in particular does not over-fit—has a test error equal to the training error. We did ten iterations of 10-fold CV, leading to 100 parameter estimates and their associated test errors.

## Modelling results I—Simultaneous fit to the data of all noise conditions

To test the null hypothesis that contrast discrimination is invariant to the presence of noise, once corrected for absolute visibility, each observer's data of both noise conditions were brought to the scale of the noiseless data (i.e., all pedestal and increment contrasts of both noise conditions were normalized by the relevant 75% correct detection threshold in noise and rescaled by the noiseless detection threshold). This operation transforms both noise detection thresholds to be equal to the noiseless detection threshold. If the invariance-hypothesis holds, this rescaling operation should also remove all systematic differences between conditions. For each observer, the four free parameter gain-control model was fitted to all rescaled data at once. An example of such a fit is shown in Figure 2a for observer L.V.

As can be seen in this figure, the TvC functions derived from the model fit are dipper shaped. Furthermore, the dependence of threshold reduction on performance level is captured by the gain-control model. At 75% correct, maximal detection threshold reduction for observer L.V. is estimated to be 2.7, at 60% correct this number equals 5, and at 90% correct it is 1.9. Figure 2b shows the distribution of deviance residuals for this fit, based on 180 blocks of 50 trials each. In Figure 2c, deviance residuals are plotted as a function of pedestal contrast with different color labels for different noise conditions.

Table 1 lists the parameter estimates, normalized total deviance, and normalized deviance split by noise condition for all observers. Parameter estimates are within the range of typical values. The normalized total deviance values, all ranging between 1.17 and 1.52, indicate a reasonable general quality of fit: The gain-control model thus explains much of the variance in these data. Nevertheless, it should be noted that for two of four observers, total deviance does not belong to the 99.9% confidence interval of the deviance expected if observers were to behave stationary. Calculating deviance by noise condition further reveals that the quality of fit differs over noise conditions: Deviance is always lowest for the moderate-noise condition but may be considerably higher for the no- and weak-noise condition (see Table 1).

To explore whether differences between noise conditions underlie this pattern, deviance residuals split by noise condition were analyzed by means of linear regression. Deviance residuals of all observers were

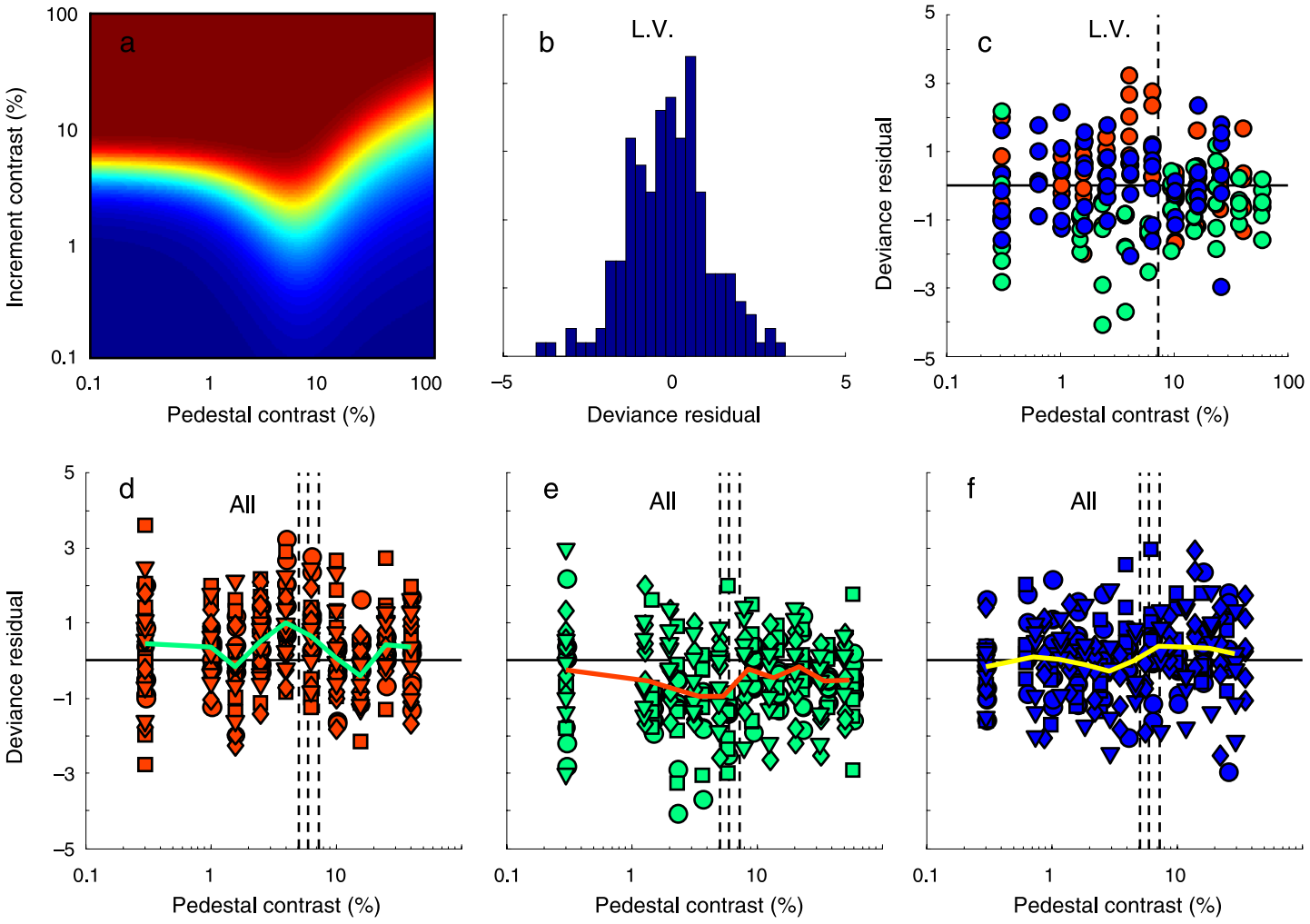


Figure 2. (a) Performance (i.e., percentage correct responses, ranging from 50%—dark blue—to 100%—dark red) of observer L.V. as a function of pedestal and increment contrast according to the best simultaneous fit of the gain-control model to all noise conditions. As can be seen, the trough of the dip is located at a pedestal contrast equal to the semisaturation contrast $\beta$. Note that psychometric functions, i.e., vertical slices, are most shallow in the trough of the dip and steepest for detection. (b) The distribution of deviance residuals for observer L.V. (c) Deviance residuals as a function of pedestal contrast for observer L.V. Different colors indicate different noise conditions (see Figure 1a). The vertical dashed line depicts $\beta$. (d–f) Deviance residuals of the simultaneous fits as a function of pedestal contrast for all observers. Different colors indicate different noise conditions; different symbols indicate different observers (see Figure 1a). The vertical dashed lines depict the $\beta$ estimates of the different observers. The thick line represents the average deviance residual at each pedestal contrast.

pooled for this analysis. Figures 2d–2f) show the deviance residuals of all observers split by noise level. The thick line describes the mean deviance residual as a function of pedestal contrast. Judged by eye, the deviance residuals of the no-noise condition (Figure 2d) seem to rise from detection threshold till around $\beta$—i.e., the semi-saturation contrast if both exponents are equal, corresponding to the location of the trough of the dip, indicated by the vertical

| | $\alpha$ | $\beta$ | $\eta$ | $\kappa$ | $\sigma$ | $D_{\text{Total}}$ | $D_{\text{NN}}$ | $D_{\text{WN}}$ | $D_{\text{MN}}$ |
|---|---|---|---|---|---|---|---|---|---|
| L.V. | 17.95 | 0.074 | 2.87 | 2.39 | **1** | 1.51** | 1.36 | 2.01** | 1.18 |
| E.G. | 10.98 | 0.050 | 3.70 | 3.19 | **1** | 1.52** | 1.88** | 1.74** | 0.93 |
| B.B. | 16.99 | 0.059 | 3.31 | 2.69 | **1** | 1.20 | 1.17 | 1.39 | 1.03 |
| L.V.E. | 10.34 | 0.051 | 3.61 | 3.14 | **1** | 1.17 | 1.13 | 1.24 | 1.14 |

Table 1. Parameter estimates and normalized deviance for the simultaneous fit of the gain-control model to all noise conditions at once. Bold symbols and numbers indicate frozen parameter values. *$D$ is outside the 99% confidence interval of the deviance of a stationary observer; **99.9% confidence interval.

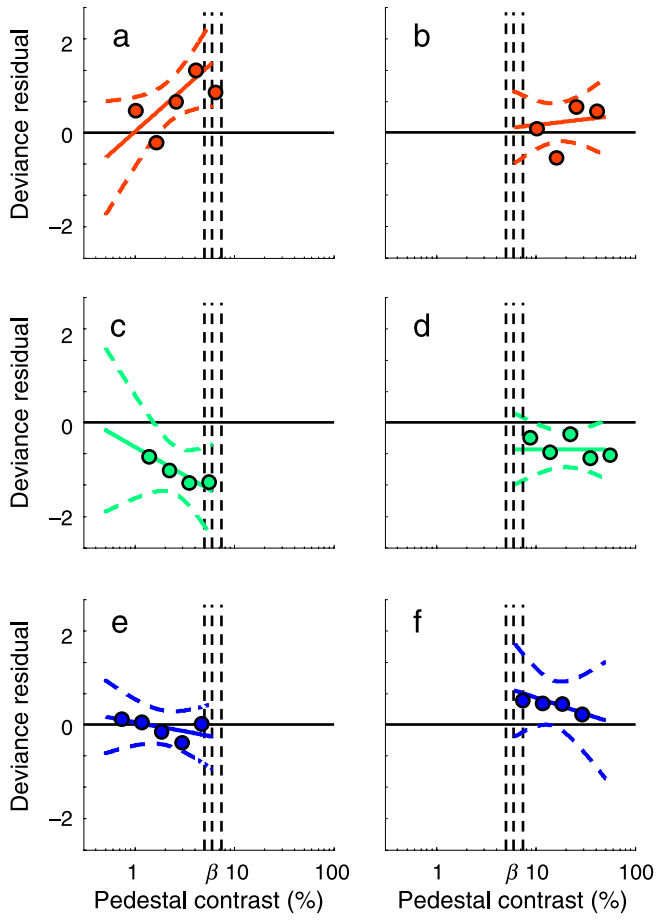


Figure 3. Results of a standard linear regression procedure, relating the logarithm of pedestal contrast to the deviance residuals of the simultaneous fit, split by noise condition. The left figure panels show the results for pedestal contrasts smaller than the semi-saturation contrast $\beta$; the right panels show the results for pedestal contrasts larger than $\beta$. Full lines indicate the best fitting linear curves; dashed lines indicate the 99.15% confidence band; symbols indicate the average deviance residual. (a–b) The no-noise condition. (c–d) The weak-noise condition. (e–f) The moderate-noise condition.

dashed lines—while the deviance residuals of the weak-noise condition (Figure 2e) display a decreasing trend in the same region. The deviance residuals of the moderate-noise condition (Figure 2f) do not display any systematic trend. Indeed, a linear regression analysis relating the logarithm of pedestal contrast to deviance residual revealed that deviance residuals of the no- and weak-noise condition differ significantly at low pedestal contrasts. Figure 3 shows the results of this linear regression analysis. The full lines depict the best fitting linear curves to the data, the dashed lines depict the 99.15% *confidence bands* of these curves, and the circles illustrate the mean deviance residual. Deviance residuals of the decreasing and rising part of the dipper function were analyzed separately. For both pedestal contrast regions, six comparisons are interesting to make: Do these three lines differ from 0, indicating a systematic misfit of the model? And do they differ from each other, indicating systematic differences between noise conditions? The overall probability of making a Type I error, i.e., falsely rejecting the null hypothesis, thus equals 0.05 (i.e., $1 - (0.9915)^6$) for both the decreasing and rising part of the dipper function.[2]

As can be seen in Figure 3, deviance residuals of the no-noise and weak-noise condition differ from zero and from each other for almost the whole decreasing part of the dipper function, while deviance residuals of the moderate noise condition are centered around zero for these pedestal contrasts. These systematic trends indicate that for low pedestal contrasts, contrast discrimination performance in the no-noise condition is underestimated by the model fit to all conditions at once, i.e., the depth of the dip is underestimated. For the weak-noise condition, on the other hand, performance, and consequently the depth of the dip, is overestimated. Making use of a computational model fitted to all data thus confirms the trend suggested by the 75% correct thresholds: Addition of both weak and moderate noise significantly—in a statistical sense—reduces the size of the pedestal effect. Furthermore, considered across all observers, the reduction seen in the presence of weak noise is more severe than in moderate noise. Modelling our data within the contrast gain-control framework thus allows us to draw statistically sound conclusions which simple confidence intervals on our raw data failed to unearth.

At higher pedestal contrasts, deviance residuals of the weak-noise condition are slightly below zero. This may indicate that weak noise affects contrast processing only at low pedestal contrasts. If this were the case, normalizing and rescaling the data by means of the detection threshold would slightly misplace the data gathered at higher contrasts. Contrast discrimination performance would be overestimated in this region, as is borne out by our data. No other differences can be noted at higher pedestal contrasts.

In sum, fitting the rescaled data of all noise conditions at once leads to a perhaps reasonable overall quality of fit, but the occurrence of systematic trends in the deviance residuals implies that the null hypothesis that contrast discrimination is invariant to the presence of noise must be rejected. Though these analyzes were run on the pooled deviance residuals, data of all observers displayed the trends described above. To analyze the differences between noise conditions in more detail, we fitted the gain-control model to the data of each noise condition separately.

## Modelling results II—Separate fits to the data of each noise condition

The most parsimonious modification of the gain-control model is to allow the response exponent and gain-control exponent to vary over noise conditions while the response

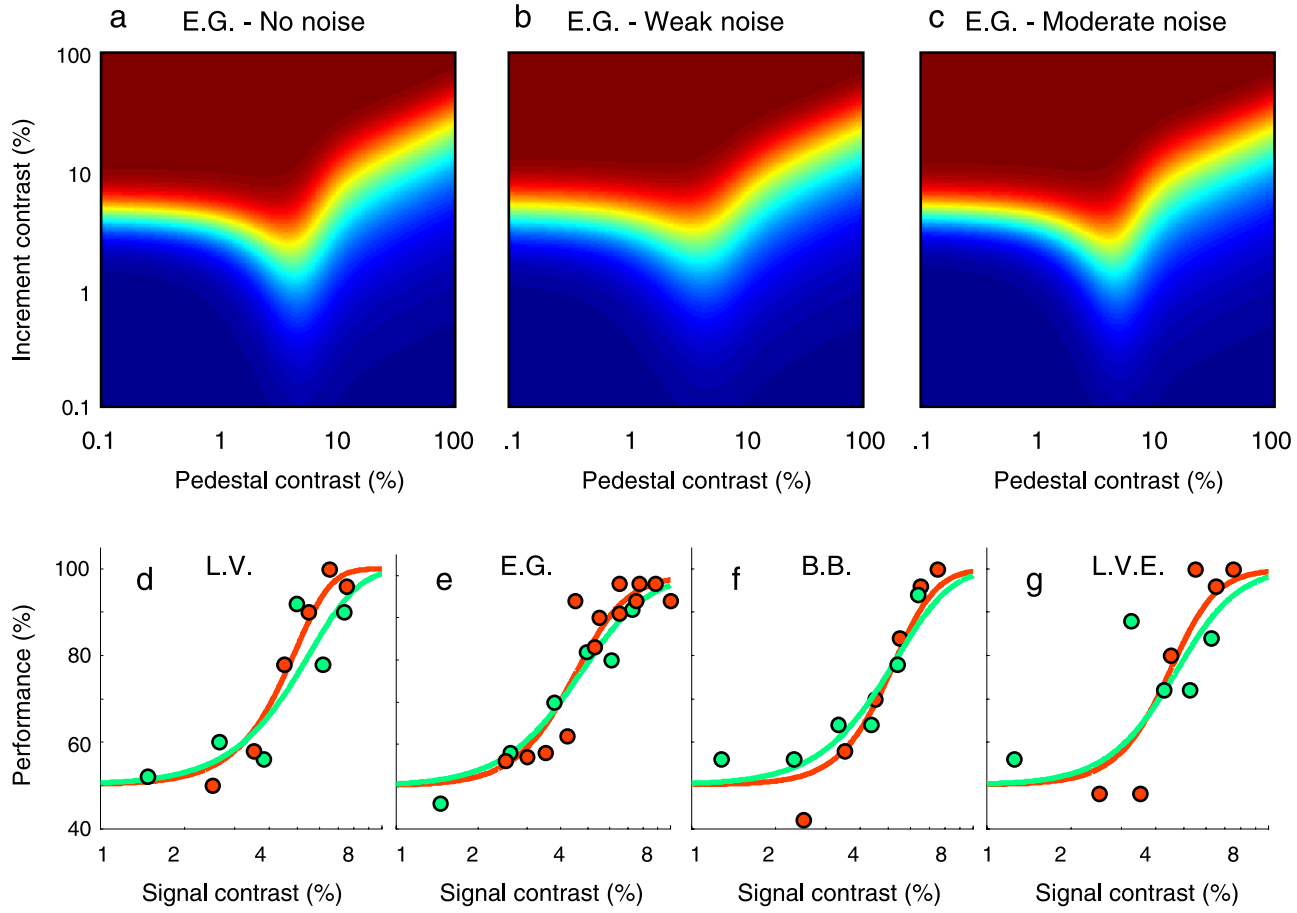


Figure 4. (a–c) Performance (i.e., percentage correct responses, ranging from 50%—dark blue—to 100%—dark red) of observer E.G. as a function of pedestal and increment contrast according to the best separate fits of the gain-control model to the different noise conditions. (d–g) The psychometric functions of all observers at zero pedestal contrast for the no-noise and weak-noise condition according to the best model fits.

gain and semisaturation contrast are frozen to the estimates of the fit to all data. Indeed, the exponents in the generalized Naka–Rushton equation determine the depth of the dip. Freezing the response gain parameter (a scaling parameter) guarantees that all models operate on the same scale and are thus easily comparable to each other and to the fit to all data at once. Freezing the semi-saturation contrast forces the dip to have the same location for all conditions, i.e., for the 75% performance contour at a contrast around the 75% correct detection threshold. In the following, the gain-control model with two effective free parameters, namely the response- and gain-control exponent $\eta$ and $\kappa$, was fitted to each noise condition separately for each observer. An example of these fits can be seen in the upper row of Figure 4 for observer E.G.

As can be seen from these plots, allowing the exponents to vary over conditions indeed leads to different estimates of the depth of the dip for different noise conditions. It is further noticeable that, in the absence of a pedestal, psychometric functions are estimated to be steeper without noise than in the presence of weak noise for each observer (see Figures 4d–4g), as was also borne out by our data (see Figure 1a). The psychometric functions derived from the model fit are plotted in the bottom row of Figure 4. Bear in mind that, due to the rescaling operation, psychometric functions are supposed to coincide at 75% correct.

Table 2 lists the parameter estimates and normalized deviance for all observers. Comparing the normalized deviance values of Table 2 with those of Table 1 reveals that quality of fit improved for each noise condition and all observers. It can be seen that normalized deviance is noticeably high for observer E.G. in the no-noise condition (i.e., $D$ = 1.87). Fitting this condition with an expanded 6-free parameter version of the gain-control model (i.e., one signal-dependent noise source, having both a multiplicative and exponential component, was added) improved quality of fit only marginally to 1.80, which is not significantly better according to AIC or BIC. Most likely, this data set is over-dispersed, i.e., observer E.G. displayed nonstationary behavior in the no-noise condition (this, of course, cannot be fixed by *any* other model: the "error" is intrinsic to the data set).

As for the model fit to all conditions at once, we analyzed the pooled deviance residuals of all observers by means of a linear regression analysis relating deviance residual to the logarithm of pedestal contrast. The raw

| | $\alpha$ | $\beta$ | $\eta$ | $\kappa$ | $\sigma$ | $D$ |
|---|---|---|---|---|---|---|
| No noise | | | | | | |
| L.V. | **17.95** | **0.074** | 3.40 | 3.00 | **1** | 1.06 |
| E.G. | **10.98** | **0.050** | 3.72 | 3.24 | **1** | 1.87** |
| B.B. | **16.99** | **0.059** | 4.17 | 3.55 | **1** | 1.00 |
| L.V.E. | **10.34** | **0.051** | 4.18 | 3.75 | **1** | 1.06 |
| Weak noise | | | | | | |
| L.V. | **17.95** | **0.074** | 2.79 | 2.22 | **1** | 1.13 |
| E.G. | **10.98** | **0.050** | 2.99 | 2.47 | **1** | 1.44 |
| B.B. | **16.99** | **0.059** | 3.06 | 2.41 | **1** | 1.37 |
| L.V.E. | **10.34** | **0.051** | 3.42 | 2.92 | **1** | 1.21 |
| Moderate noise | | | | | | |
| L.V. | **17.95** | **0.074** | 2.68 | 2.23 | **1** | 1.16 |
| E.G. | **10.98** | **0.050** | 3.91 | 3.39 | **1** | 0.93 |
| B.B. | **16.99** | **0.059** | 2.94 | 2.33 | **1** | 1.00 |
| L.V.E. | **10.34** | **0.051** | 3.17 | 2.70 | **1** | 1.11 |

Table 2. Parameter estimates and normalized deviance for the fit of the gain-control model to the separate noise conditions. Bold symbols and numbers indicate frozen parameter values. *$D$ is outside the 99% confidence interval of the deviance of a stationary observer; **99.9% confidence interval.

data of this analysis are plotted in Figure 5, the summary of this analysis in Figure 6. For the decreasing part of the dipper, all differences between noise conditions have vanished. As we hypothesized, allowing the response- and gain-control exponent to vary over noise conditions is sufficient to capture the systematic differences between conditions at low pedestal contrasts. At higher contrasts, virtually nothing has changed, so there is still a small but systematic misfit for the weak-noise condition.

In sum, freezing the response gain ($\alpha$) and semisaturation contrast ($\beta$) to the estimates of the fit to the pooled noise conditions and leaving the exponents ($\eta$ and $\kappa$) free to capture the differences between the noise conditions leads to a parsimonious model (i.e., 0.033 free parameters per block of 50 trials which corresponds to about 1 free parameter per 1,500 trials) that successfully describes our data. Compared to the simultaneous fit to all data, allowing the exponents to vary over noise conditions leads to an improvement in quality of fit and the disappearance of systematic trends in the deviance residuals at low pedestal contrasts. We now assess whether the improvement brought about by more free parameters is sufficiently large as assessed by methods of model selection.

## Model selection: Simultaneous vs. separate fits

As explained in the paragraph on model selection, there are different ways to assess which of the two modelling approaches has the highest predictive accuracy. We first consider the outcome of the AIC procedure in detail. First, and most general, we may consider the overall predictive accuracy for all observers and all conditions, analog to pooling of data across observers and conditions. We must thus compare the AIC of a 16-free parameter model (i.e., $4 \times 4$ free parameters) to the AIC of a 24-free parameter model (i.e., $4 \times 3 \times 2$ adjustable parameters) for a data set consisting of 727 blocks. For the fit to all conditions at once, AIC is 992.92; for the separate fits, AIC is 893.19. The reduction in AIC thus equals 99.73. This value is notably higher than any sensible critical value derived from a $\chi^2$-distribution with eight free parameters (e.g., 99.9% of the area of this distribution is located below 26.125). We may thus conclude that, considered across noise conditions and observers, the response gain and semisaturation contrast may be frozen, but it is better not to freeze the exponents of the Naka–Rushton equation.

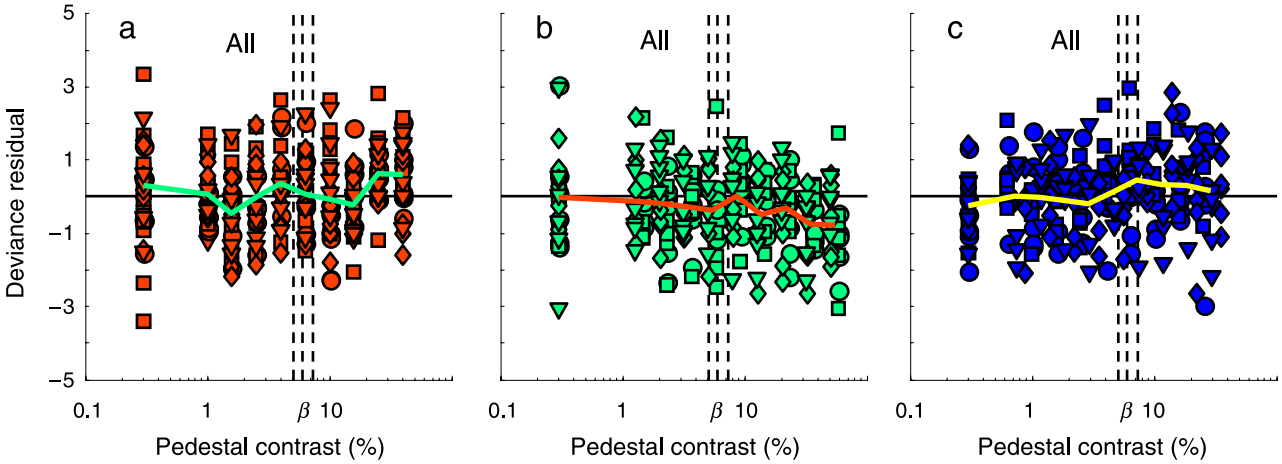


Figure 5. Deviance residuals of the separate fits as a function of pedestal contrast for all observers. Different colors indicate different noise conditions; different symbols indicate different observers. The vertical dashed lines depict the $\beta$-estimates of the different observers. The thick line represents the average deviance residual.

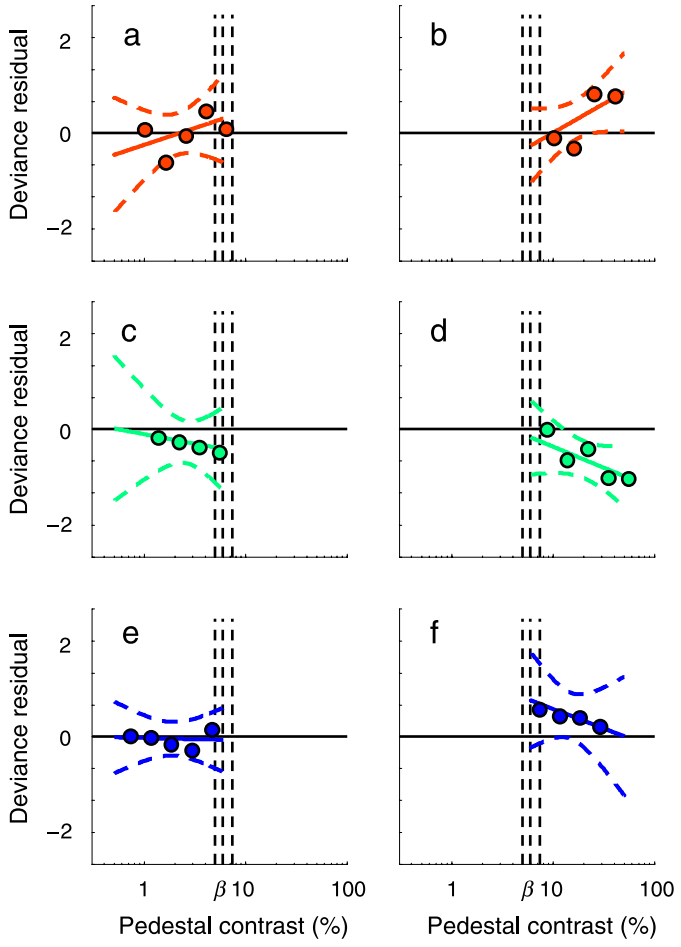


Figure 6. Results of a standard linear regression procedure, relating the logarithm of pedestal contrast to the deviance residuals of the separate fits, split by noise condition. The left figure panels show the results for pedestal contrasts smaller than the semi-saturation contrast $\beta$; the right panels show the results for pedestal contrasts larger than $\beta$. Full lines indicate the best fitting linear curves; dashed lines indicate the 99.15% confidence band; symbols indicate the average deviance residual. (a–b) The no-noise condition. (c–d) The weak-noise condition. (e–f) The moderate-noise condition.

| | No noise | Weak noise | Moderate noise | ∑ |
|---|---|---|---|---|
| L.V. | 16.78*** | 51.01*** | 0.36 | 68.15*** |
| E.G. | −2.16 | 17.27** | −0.80 | 14.31*** |
| B.B. | 9.45** | 1.14 | 1.02 | 11.61** |
| L.V.E. | 3.37 | 1.25 | 1.04 | 5.66 |
| ∑ | 27.44*** | 70.67*** | 1.62 | |

Table 3. Reduction in AIC for each observer and each noise condition. Significant reduction at *$\alpha$ = 0.05, **$\alpha$ = 0.005, and ***$\alpha$ = 0.001. Whenever the theoretical degrees of freedom were not a natural number (e.g., 2.66), we used the conservative critical value of the $\chi^2$ distribution of the next natural number (i.e., 3).

| | No noise | Weak noise | Moderate noise | ∑ |
|---|---|---|---|---|
| L.V. | 15.50 | 49.73 | −0.91 | 64.32 |
| E.G. | −3.44 | 16.00 | −2.07 | 10.49 |
| B.B. | 8.18 | −0.13 | −0.25 | 7.8 |
| L.V.E. | 2.56 | 0.44 | 0.23 | 3.23 |
| ∑ | 22.81 | 66.04 | −3 | |

Table 4. Reduction in BIC for each observer and each noise condition.

Second, we can also do this analysis for each noise condition across observers and for each observer across noise conditions. And finally, we can do this analysis for each condition within each observer.

Results of these analyses are summarized in Table 3. Each cell denotes an analysis at the third, most detailed level. Each marginal total denotes an analysis at the second, intermediate level of detail. Significant reductions in AIC are marked by means of stars in Table 3. It will be noted that the second model is favored over the first for three of four observers. For observer L.V.E., there is a trend in the same direction, but the reduction in AIC is only marginally significant ($AIC_1 - AIC_2 = 5.66$; $p < 0.059$). We suspect that this may at least in part result from the relative lack of data that reduces statistical power —she completed only 4,500 trials—because her deviance residuals and parameter estimates are not inconsistent with other observers. When considering the different noise conditions, it is clear that the improvement in predictive accuracy is mainly due to the better fits to the weak-noise condition ($AIC_1 - AIC_2 = 70.67$; $p < 10^{-5}$) and the no-noise condition ($AIC_1 - AIC_2 = 27.44$; $p < 10^{-5}$). For the moderate-noise condition, fits were already fine in the first approach, so not much could be gained ($AIC_1 - AIC_2 = 1.62$; $p < 0.65$).

Table 4 summarizes the reduction in BIC. Positive numbers indicate that the separate fits should be selected, negative numbers that the simultaneous fit should be selected. As can be seen in Table 4, conclusions of the BIC analysis mostly agree with the AIC analysis. At the most general level, i.e., considering the pooled data of all

| | Model I | Model II | Model I–Model II |
|---|---|---|---|
| L.V. | 1.54** | 1.16 | 0.39 |
| E.G. | 1.54** | 1.47** | 0.07 |
| B.B. | 1.24* | 1.15 | 0.09 |
| L.V.E. | 1.19 | 1.16 | 0.04 |
| ∑ | 1.38 | 1.24 | 0.15 |

Table 5. Average test error for each observer across noise conditions for the simultaneous and separate fits and their difference. *Average $D_{Test}$ is outside the 99% confidence interval of the deviance of a stationary observer; **99.9% confidence interval. For Model II, the presence of stars indicates that the average $D_{Test}$ was outside the relevant confidence interval for *at least one* noise condition.

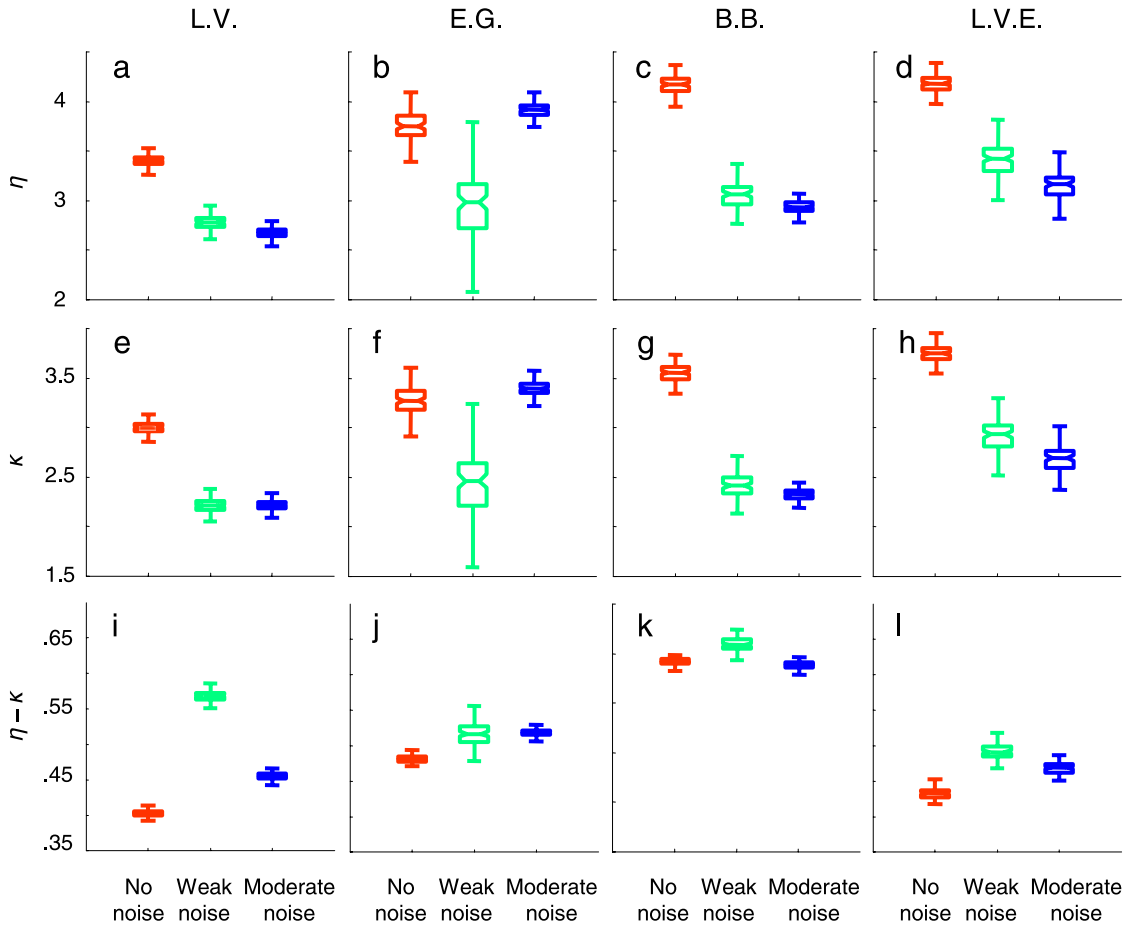


Figure 7. (a–d) Box plots of the 100 $\eta$-values, i.e., the response exponent, estimated in the cross-validation for all observers. (e–h) Box plots of the 100 $\kappa$-values, i.e., the gain-control exponent, estimated in the cross-validation for all observers. (i–l) Box plots of the 100 differences between $\eta$ and $\kappa$ estimated in the cross-validation analysis for all observers. In all these box plots, the central horizontal line indicates the second quartile (i.e., the median), while the other horizontal lines indicate the first and third quartile. Notches around the median indicate the 95% confidence interval of the median. Whiskers indicate one and a half times the interquartile range. Outliers have been omitted for clarity.

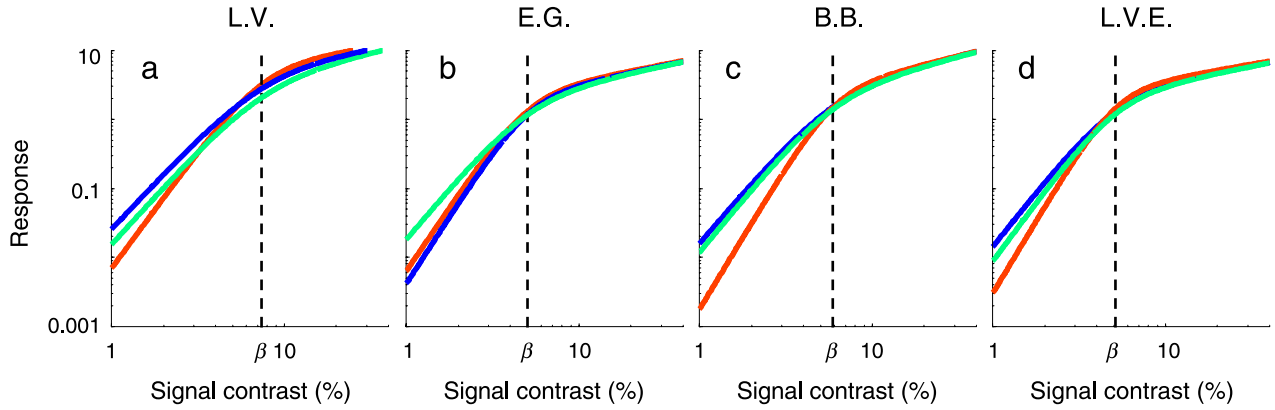


Figure 8. Contrast response functions according to the best separate fits of the gain-control model to the different noise conditions, plotted on double logarithmic coordinates. Red lines refer to the no-noise condition, green to the weak-noise condition, and blue to the moderate-noise condition.

observers and noise conditions, the second modelling approach does much better than the first, according to BIC. At the level of individual observers, the same trend is obvious. At the level of noise conditions, the no-noise and weak-noise condition clearly benefit from the additional free parameters in the separate fits. This is not the case for the moderate noise condition, for which BIC selects the simultaneous fit.

Table 5 summarizes the average CV index (i.e., *test error*) and the difference in test error for both modelling approaches for each observer. As indicated by the positive differences in test error, the separate fits provide better predictions for *unseen* data for each observer. It should in addition be noted that for three of four observers, the average test error of each noise condition belongs to the 99% confidence interval of the distribution of deviance values of a stationary observer: The data are thus very well described by our model.

In sum, despite differences in the aspects of model complexity captured by AIC, BIC, and CV, all support the same conclusion: When pooled data and observers are considered, the separate fits are always selected over the joint fit. This conclusion also holds at the level of individual observers. At the level of the noise conditions, the improved predictive accuracy is a result of the better fit to both the no-noise and weak-noise condition.

## Discussion

Given our comprehensive model selection, we are now in a position to use the parameter estimates of the cross-validation to assess differences between noise conditions. Because the response gain and semisaturation contrast

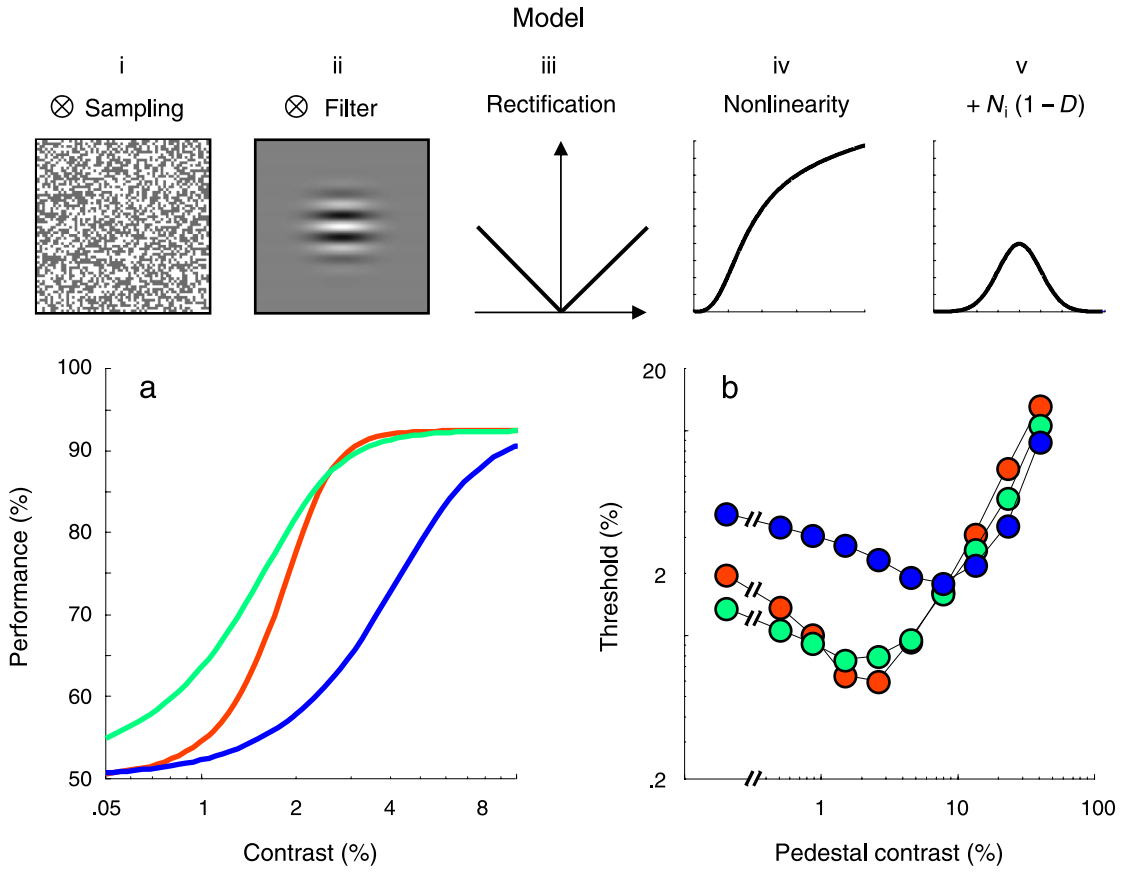


Figure 9. (Upper row) Illustration of the model components used in the simulation. Stimuli with a narrow-band spatial-frequency spectrum are often assumed to be processed in a single spatially localized, spatial-frequency, and orientation-selective filter. Ideally, an observer uses a filter that is an exact template of the signal (ii). Human deviations from ideal observer performance are typically interpreted as stemming from internal noise sources (v) on the one hand and limited efficiency or sampling (i) on the other hand (this refers to using only a sample of the available image information, for instance only the image pixels corresponding to the white pixels in i). The transduction mechanism further consisted of rectification (iii) and a nonlinear post-filter stage—the generalized Naka–Rushton transformation (iv). The final model component was a late, signal-independent noise source (v). (Lower row; a) Simulated performance as a function of signal contrast on semi-logarithmic coordinates for the three detection conditions. Red lines refer to the no-noise condition, green to the weak-noise condition, and blue to the moderate-noise condition. (b) The simulated TvC functions at 75% correct.

were frozen, we only need to consider the exponents of the Naka–Rushton function. Box plots of the estimates of the response exponent $\eta$, the gain-control exponent $\kappa$, and their difference are shown in Figure 7. We first compare parameter estimates for the no-noise and weak-noise condition. For all observers, both exponents are estimated to be reduced in the presence of weak noise. Furthermore, the difference between the exponents is always higher for the weak-noise condition than for the no-noise condition. It is this difference, together with the absolute value of the response exponent, which determines the strength of the pedestal effect: The smaller the difference and the larger the response exponent, the bigger the pedestal effect is. In other words, the depth of the dipper function is reduced in the presence of weak noise for all observers. It is interesting to note that results are not as systematic for the moderate-noise condition. For some observers, exponents are estimated to be reduced relative to the no-noise condition (e.g., observer B.B.), but for others, this is clearly not the case (e.g., observer E.G.). A similar variability over observers is present in the differences between both exponents: For some this difference has increased in the presence of moderate noise (e.g., observer L.V.E.), but for others this is not the case (e.g., observer B.B.). This variation over observers is not inconsistent with data sets that have been published earlier: The dipper function of some observers seems to be invariant to the presence of strong noise, while this is not the case for others (e.g., Henning & Wichmann, 2007).

We thus find three stable differences between the no noise and weak-noise condition, namely, a reduction in both the response- and gain-control exponent and an increased difference between these exponents in the presence of weak noise. This indicates that the pedestal effect was reduced for all observers in the presence of weak noise. For the moderate-noise condition, results vary over observers.

To better understand these differences between noise conditions, it is helpful to "open" the models and have a look at the internal contrast response functions (see Figure 8; cf. Kienzle, Wichmann, Schölkopf, & Franz, 2007; Wichmann, Graf, Simoncelli, Bülthoff, & Schölkopf, 2005). Plotted on double logarithmic coordinates, internal response rises steeply until $\beta$, the semi-saturation contrast. In this region, the contrast response behaves as an accelerating nonlinearity. Because internal variance is constant at all contrast levels in our fits ($\sigma = 1$), these functions could also be interpreted as detection functions (signal-to-noise ratio as a function of contrast). Indeed, detection sensitivity has been reported to rise in an accelerating way as a function of contrast (Foley & Legge, 1981; Nachmias, 1981; Nachmias & Sansbury, 1974). It is this acceleration that leads to the response expansion that underlies the pedestal effect. The larger the log–log steepness at low contrasts, the stronger the pedestal effect. Comparing the no-noise condition to the weak-noise condition illustrates that the log–log steepness, and thus the pedestal effect, is reduced for all observers in the presence of weak noise. The reduced log–log steepness in weak noise is a consequence of the higher response at low contrasts, which in turn leads to improved sensitivity. The response difference between the no-noise and weak-noise condition diminishes as a function of contrast and disappears completely around the semi-saturation contrast (due to the rescaling procedure). This is not inconsistent with the effect of a rectification mechanism at the output of a linear filter stage prior to the nonlinear response expansion. Due to rectification, the mean response of a linear filter is enhanced in the presence of noise. At zero pedestal contrast, the response-enhancing effect of rectification is maximal because half of the responses of a nonrectified linear filter are negative. As pedestal contrast, and thus the average filter response, increases, the proportion of negative responses drops and the response-enhancing effect of rectification diminishes until it vanishes completely. This is consistent with the contrast response functions plotted in Figure 8.

Thus, using the gain-control model to get a small parameter description of our data that is statistically sound indeed shows that some parameters change in the presence of noise. We do not wish to claim that these parameters actually change in the visual system, but rather that this is the statistical signature of how noise changes the data.

To further explore whether one mechanism may underlie the data, we simulated performance of the Goris et al. (2008) gain-control model in the contrast discrimination experiments reported here (see Figure 9, upper row). In this simulation, the filter stage—which need not be specified explicitly for sinusoidal contrast discrimination—was chosen to consist of *(optimal) template matching,* followed by full-wave rectification (e.g., Lu & Dosher, 2008). Template matching is a convenient way to transform 2-D input images to 1-D "filter responses." Prior to this filter stage, *stimulus sampling* or limited calculation efficiency was assumed, as is often the case in detection-in-noise models (e.g., Lu & Dosher, 2008). This is described by parameter $k$, which expresses the proportion of available information used by the observer and ranges between 0 and 1. Inefficiencies in the visual system need not be conceptualized as sampling: parameter $k$ could be thought of as, e.g., reflecting the use of a suboptimal filter, for instance a spatial-frequency tuned channel that has an effective bandwidth that is broader than the narrowband Gabor signal. To describe the nonlinear mapping of stimulus contrast to internal contrast representation, the second part of the transduction mechanism consisted of the generalized four parameter Naka–Rushton function (free parameters $\alpha$, $\beta$, $\eta$, and $\kappa$). The rectified filter responses used in the expansive, i.e., the numerator, and the compressive, i.e., the denominator, parts of the Naka–Rushton function were the same. Although some evidence points to the existence of a

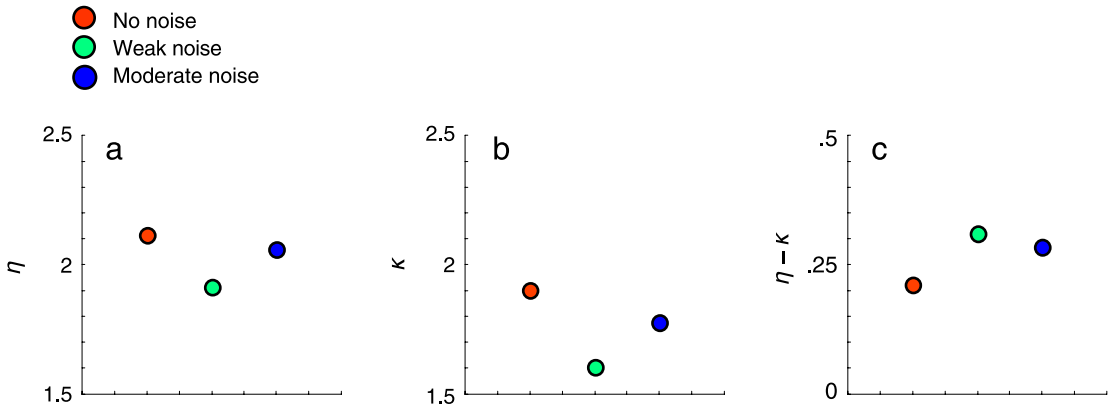


Figure 10. (a) The response exponent estimates for the simulated data set. (b) The gain-control exponent estimates for the simulated data. (c) The differences between $\eta$ and $\kappa$ for the simulated data set.

broadly tuned contrast gain-control pool (e.g., Foley, 1994; Holmes & Meese, 2004), we opted to use only within-channel suppression in this model to avoid an increase of the number of parameters. Furthermore, because the spectral properties of the noise were constant for the weak and moderate noise conditions, it is unlikely that this simplification has a significant impact on our conclusions. The transduction mechanism in this simulation is thus fully determined by specifying the sampling ($k$) and the parameters of the generalized Naka–Rushton equation ($\alpha$, $\beta$, $\eta$, and $\kappa$).

The variability of internal states is determined by internal noise on the one hand and the use of stochastic stimuli on the other hand. One parameter was used to describe the late 1-D internal noise, assumed to be Gaussian and signal-independent (parameter $\sigma$). The effect of the external noise on internal variability was derived from Monte Carlo simulations. With the noise levels used in our experiments as input to the gain-control model, $k$, $\eta$, $\kappa$, and $\beta$ were varied—effects of $\alpha$, i.e., the response gain, need not be simulated. Descriptive functions were fitted to the simulated response variances. Assuming equal variance of the signal-plus-noise and noise representations (as a first approximation), allowed us to use these descriptive functions to formalise the full model behavior.

In sum, the model used in this simulation consisted of *image sampling* ($k$), *template matching, response rectification, nonlinear transduction* ($\alpha$, $\beta$, $\eta$, and $\kappa$), and *late noise addition* ($\sigma$). All these components are illustrated in the upper row of Figure 9. As we did in the rest of the paper, $\sigma$ was frozen to 1. The other five model parameters were chosen in such a way that the (normalized) simulated thresholds would approximate the data shown in Figure 1. We found this to be the case for $k = 0.05$ (i.e., sampling), $\alpha = 2*10^6$ (i.e., the response gain), $\beta = 0.025$ (i.e., the semisaturation contrast), $\eta = 3$ (i.e., the response exponent), $\kappa = 1.75$ (i.e., the gain-control exponent), and $\sigma = 1$ (i.e., the late noise). Figure 9a shows the psychometric functions relating detection performance to signal contrast without noise (red) and in the presence of weak (green) and moderate noise (blue) on semi-logarithmic coordinates. As was the case for the detection in noise data discussed in this paper and shown in Figure 1a, addition of weak noise improves contrast detection performance. This is not the case for moderate noise. In weak noise, the 75% correct detection threshold is reduced by a factor of 1.39 (as compared to 1.39 across observers). In modest noise, this threshold is increased by a factor of 2.14 (1.34 across observers). Further, psychometric functions are not parallel on semi-logarithmic coordinates. With these parameter values, the Goris et al. (2008) gain-control model thus mimics some aspects of human detection in noise data.

Figure 9b shows the full simulated TvC functions at 75% correct for all noise conditions on double logarithmic coordinates (compare to Figures 1b and 1c). The most leftward points denote detection (i.e., the pedestal contrast equals zero). For these TvC functions, maximal threshold reduction at 75% correct was of a factor 3.38 without noise (3.44 across observers), 1.78 in the presence of weak noise (2.17 across observers), and 2.23 in the presence of moderate noise (2.62 across observers). As was the case for the discrimination data discussed in this paper and shown in Figures 1b and 1c, the depth of the dip is thus most reduced in the presence of weak noise. It may further be noticed that contrast discrimination thresholds of the three conditions almost coincide at high pedestal contrasts, consistent with our data. At these contrasts, the slope of the dipper handle approximates one on double logarithmic coordinates, in line with several psychophysical observations (e.g., Bird et al., 2002). In summary, a single contrast processing mechanism can produce contrast discrimination in noise data that resemble many aspects of our experimental findings.

Finally, as we did for our observers, the gain-control model used throughout the paper was also fitted to the normalized and rescaled simulated data, split per noise

level (in line with the previous model fitting, the response gain and semisaturation contrast were frozen to the estimates of the fit to all data, i.e., $\alpha = 12.3$ and $\beta = 0.04$). Note that this gain-control model is similar but not identical to the model used to produce these data. The most interesting parameter estimates are shown in Figure 10 (compare to Figure 7). It will be noted that the three parameter changes discussed above are also present in the parameter estimates for the simulated data set (i.e., a reduction in both the response- and gain-control exponent and an increased difference between these exponents in the presence of weak noise, relative to no-noise). This simulation thus shows that a single mechanism underlying the pedestal effect and stochastic resonance may have the signature of reduced exponents if the gain-control model is fitted to the data. These similarities further support the interpretation of the contrast discrimination data discussed in this paper as consistent with the idea that a single mechanism underlies the pedestal effect in contrast discrimination and stochastic resonance in contrast detection.

It is harder to explain how contrast processing changes in the presence of moderate noise. A main part of the problem is that data differ considerably between observers. For some observers, the pedestal effect is much reduced in moderate noise. But for others, hardly anything seems to differ. An easy, common interpretation is thus unlikely to be found.

## Conclusion

In this paper we explored whether a single mechanism underlies the pedestal effect, i.e., the improved detectability of a grating in the presence of a low-contrast masking grating, and stochastic resonance, i.e., the improved detectability of a grating in the presence of subthreshold noise. Analysis of the 75% correct thresholds was not conclusive, i.e., the trends present in the data were not significant. Making use of a full quantitative description of our data with few parameters, i.e., using the gain-control model as a statistical tool, combined with comprehensive model selection assessments, we showed the pedestal effect to be *reduced* in the presence of weak noise for all observers. This reduction clearly rules out independent, additive sources of performance improvement and cannot simply be attributed to additionally introduced response variability by the weak noise because it was smaller and not as consistent in the presence of moderate noise. We further showed that a single mechanism responsible for the pedestal effect and stochastic resonance may have the signature of reduced exponents if the gain-control model is fitted to the data. Given that the pattern of parameter changes for real data is the same as for simulated data (under the hypothesis of a single mechanism) and that the alternative hypothesis can be ruled out by model selection, we interpret these data as indicating that a single mechanism underlies the pedestal effect and stochastic resonance in contrast perception.

## Acknowledgments

We would like to thank G. B. Henning and two anonymous reviewers for their helpful comments. In particular, we are grateful to reviewer two who suggested we should add a set of simulations showing how the parameters of our descriptive model change for simulated data. R.G. is research assistant of the Fund for Scientific Research-Flanders (FWO-Vlaanderen) under the supervision of J.W. and F.A.W.

Commercial relationships: none.
Corresponding author: Robbe L. T. Goris.
Email: Robbe.Goris@psy.kuleuven.be.
Address: Laboratory for Experimental Psychology, Tiensestraat 102, B-3000 Leuven, Belgium.

## Footnotes

[1]Note that this is not as straightforward as it may sound, as functions with nominally the same number of free parameters may have inherently more or less complexity. For a nontechnical introduction, see Forster (1999, 2000) and Forster and Sober (1994).

[2]Detection data were omitted from the low-pedestal contrast analysis. The reasons are twofold. First, due to the normalisation procedure, most differences between deviance residuals as a function of pedestal contrast will be removed for detection. Second, because of the logarithmic transformation of signal contrast, there is no correct location for these deviance residuals on the contrast axis.

## References

Bach, M. (1997). A note on luminance calibration of raster-scan cathode-ray tubes: Temporal resolution, ripple, and accuracy. *Spatial Vision, 10,* 485–489. [PubMed]

Bach, M., Meigen, T., & Strasburger, H. (1997). Raster-scan cathode-ray tubes for vision research–limits of resolution in space, time and intensity, and some solutions. *Spatial Vision, 10,* 403–414. [PubMed]

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

Bird, C. M., Henning, G. B., & Wichmann, F. A. (2002). Contrast discrimination with sinusoidal gratings of different spatial frequency. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 19,* 1267–1273. [PubMed]

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

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. *The Journal of Physiology, 203,* 237–260. [PubMed] [Article]

Bradley, A., & Ohzawa, I. (1986). A comparison of contrast detection and discrimination. *Vision Research, 26,* 991–997. [PubMed]

Brainard, D. H. (1989). Calibration of a computer controlled colour monitor. *Color Research and Application, 14,* 23–34.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology, 44,* 108–132. [PubMed]

Campbell, F. W., Carpenter, R. H., & Levinson, J. Z. (1969). Visibility of aperiodic patterns compared with that of sinusoidal gratings. *The Journal of Physiology, 204,* 283–298. [PubMed] [Article]

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology, 197,* 551–556. [PubMed] [Article]

Collins, J. J., Imhoff, T. T., & Grigg, P. (1996). Noise-enhanced tactile sensation. *Nature, 383,* 770. [PubMed]

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A, Optics and Image Science, 2,* 1160–1169. [PubMed]

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

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

Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 11,* 1710–1719. [PubMed]

Foley, J. M., & Chen, C. C. (1997). Analysis of the effect of pattern adaptation on pattern pedestal effect: A two-process model. *Vision Research, 37,* 2779–2788. [PubMed]

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

Foley, J. M., & Schwarz, W. (1998). Spatial attention: Effect of position uncertainty and number of distracter patterns on the threshold-versus-contrast function for contrast discrimination. *Journal of the Optical Society of America A, Optics and Image Science, 15,* 1036–1047.

Forster, M. R. (1999). Model selection in science: The problem of language variance. *British Journal for the Philosophy of Science, 50,* 83–102.

Forster, M. R. (2000). Key concepts in model selection: Performance and generalizability. *Journal of Mathematical Psychology, 44,* 205–231. [PubMed]

Forster, M. R., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science, 45,* 1–35.

Georgeson, M. A., & Georgeson, J. M. (1987). Facilitation and masking of briefly presented gratings: Time-course and contrast dependence. *Vision Research, 27,* 369–379. [PubMed]

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

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

Goris, R. L. T., Zaenen, P., & Wagemans, J. (2008). Some observations on contrast detection in noise. *Journal of Vision, 8*(9):4, 1–15, http://journalofvision.org/8/9/4/, doi:10.1167/8.9.4. [Article]

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

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley & Sons, Inc.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2003). *The elements of statistical learning theory*. New York: Springer.

Henning, G. B. (1988). Spatial-frequency tuning as a function of temporal frequency and stimulus motion. *Journal of*

*the Optical Society of America A, Optics and Image Science, 5,* 1362–1373. [PubMed]

Henning, G. B., Bird, C. M., & Wichmann, F. A. (2002). Contrast discrimination with pulse trains in pink noise. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 19,* 1259–1266. [PubMed]

Henning, G. B., Derrington, A. M., & Madden, B. C. (1983). Detectability of several ideal spatial patterns. *Journal of the Optical Society of America, 73,* 851–854. [PubMed]

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

Henning, G. B., Hertz, B. G., & Hinton, J. L. (1981). 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. [PubMed]

Henning, G. B., & Wichmann, F. A. (2007). Some observations on the pedestal effect. *Journal of Vision, 7*(1):3, 1–15, http://journalofvision.org/7/1/3/, doi:10.1167/7.1.3. [PubMed] [Article]

Holmes, D. J., & Meese, T. S. (2004). Grating and plaid masks indicate linear summation in a contrast gain pool. *Journal of Vision, 4*(12):7, 1080–1089, http://journalofvision.org/4/12/7/, doi:10.1167/4.12.7. [PubMed] [Article]

Kienzle, W., Wichmann, F. A., Schölkopf, B., & Franz, M. O. (2007). A nonparametric approach to bottom-up visual saliency. In B. Schölkopf, J. Platt, and T. Hoffman (Eds.), *Advances in neural information processing systems* (vol. 19, pp. 689–696). Cambridge, MA: MIT Press.

Kontsevich, L. L., Chen, C. C., & Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Research, 42,* 1771–1784. [PubMed]

Legge, G. E. (1981). A power law for contrast discrimination. *Vision Research, 21,* 457–467. [PubMed]

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

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A, Optics and Image Science, 4,* 391–404. [PubMed]

Lu, Z. L., & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review, 115,* 44–82. [PubMed]

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44,* 190–204. [PubMed]

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

Nachmias, J., & Sansbury, R. V. (1974). Letter: Grating contrast: Discrimination may be better than detection. *Vision Research, 14,* 1039–1042. [PubMed]

Naiman, A. C., & Makous, W. L. (1992). Spatial nonlinearities of grayscale CRT pixels. In *Human vision, visual processing, and digital display III, SPIE Proceedings 1666* (pp. 41–56). Bellingham, WA: SPIE.

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

Pelli, D. G. (1997a). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–42. [PubMed]

Pelli, D. G. (1997b). Pixel independence: Measuring spatial interactions on a CRT display. *Spatial Vision, 10,* 443–446. [PubMed]

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6,* 421–425. [PubMed]

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472–491. [PubMed]

Smithson, H. E., Henning, G. B., MacLeod, D. I. A., & Stockman, A. (in press). The effect of notched noise on flicker detection and discrimination.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44,* 92–107. [PubMed]

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

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

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

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and

sampling. *Perception & Psychophysics, 63,* 1314–1329. [PubMed]

Wichmann, F. A., Graf, A. B. A., Simoncelli, E. P., Bülthoff, H. H., & Schölkopf, B. (2005). Machine learning applied to perception: Decision-images for gender classification. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing system* (Vol. 17, pp. 1489–1496). Cambridge, MA: MIT Press.

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

Yang, J., & Makous, W. (1995). Modeling pedestal experiments with amplitude instead of contrast. *Vision Research, 35,* 1979–1989. [PubMed]

Yu, C., Klein, S. A., & Levi, D. M. (2002). Facilitation of contrast detection by cross-oriented surround stimuli and its psychophysical mechanisms. *Journal of Vision, 2*(3):4, 243–255, http://journalofvision.org/2/3/4/, doi:10.1167/2.3.4. [PubMed] [Article]

Zeng, F. G., Fu, Q. J., & Morse, R. (2000). Human hearing enhanced by noise. *Brain Research, 869,* 251–255. [PubMed]

Zucchini, W. (2000). An introduction to model selection. *Journal of Mathematical Psychology, 44,* 41–61. [PubMed]