Labeled lines for image blur and contrast

Michael J. Morgan

Division of Optometry and Visual Sciences, City, University of London, London, UK
Max-Planck Institute for Metabolism Research, Cologne, Germany

It has been suggested that blur and contrast discrimination thresholds are limited by a common stage of contrast energy transduction, and that this explains the characteristic "dipper" functions found for contrast and blur discrimination. To test this conjecture, thresholds for discriminating increments from decrements in sharpness/blur, and similarly for contrast, were measured using the same chessboard stimuli with the Method of Single Stimuli (Experiment 1). Using a generic human contrast sensitivity function (HCSF) to calculate energy, thresholds were significantly lower for blur than for contrast. They could be made more similar only by using an implausibly narrow band-pass version of the HCSF. In separate sessions (Experiment 2), observers also attempted to discriminate between blur and contrast changes when they were randomly interleaved (channel discrimination). Channel discrimination thresholds were similar to those predicted from noisy independent channels, consistent with separate labeled lines for the two channels. Experiment 3 measured subthreshold summation of contrast and blur signals, in either energy-add or energy-subtract modes with a two-alternative forced choice task. Both add and subtract modes lowered thresholds. Experiment 4 measured standard T versus C ("dipper") functions for blur, and compared these with T versus C functions when a contrast cue was added to keep energy constant. The finding of a "dipper" function in the latter case suggests that it does not arise from a common energy transduction stage.

Introduction

Blur is the informal description of what happens to an image when it is visibly not in focus. All images from familiar optical instruments such as cameras and telescopes are blurred because all of these imaging devices spread a single point of light, however small, into a dispersed *point spread function* in the image. *Contrast* is the informal description of the range of black and white values in the image. Ordinary observers without optical training are perfectly able to pick out which of a range of images has the highest blur, or the highest contrast.

Because blur and contrast are both fundamental physical properties of images, it is natural to ask how they are physically and psychophysically related. Informally again, an interaction between blur and contrast is indicated by the fact that it is hard to see whether a very low contrast image is blurred or not. This could be because both blur and low contrasts remove high spatial frequencies from the "window of visibility" (Watson, Ahumada, & Farrell, 1986). Watson and Ahumada (2011) provide a valuable review of blur discrimination experiments and find that they are in broad empirical agreement. However, contrary to previous modeling efforts, they conclude that specialized mechanisms are not required and that the essential features of blur discrimination are fully accounted for by a visible contrast energy (ViCE) model, in which two spatial patterns are distinguished when the contrast energy of their filtered difference reaches a threshold value. In the ViCE model, intrinsic blur is represented by the high-frequency limb of the contrast sensitivity function, but the low-frequency limb also contributes to the predictions for large reference blurs, and the model includes masking, which improves predictions for high-contrast stimuli.

The Watson–Ahumada model is unquestionably an advance on previous models because it provides a common metric (visible contrast energy) by means of which different tasks can be compared. Ideally, energy thresholds should be the same for contrast and blur discrimination, but it is clear from the data analyzed by Watson and Ahumada that those for blur are generally lower, particularly at low pedestal values. However, there are no experiments of which we are aware in which contrast and blur discrimination have been measured with identical stimuli in the same observers. Watson and Ahumada were forced to rely on comparison between different experiments, some of which attracted their justified disapproval. The main purpose of the experiments reported in this paper is to place comparisons of blur and contrast discrimination on a sound footing, by using identical stimuli, backgrounds, and procedures in the two cases.

Citation: Morgan, M. J. (2017). Labeled lines for image blur and contrast. *Journal of Vision*, *17*(6):16, 1–11, doi:10.1167/17.6.16.

doi: 10.1167/17.6.16 Received July 29, 2016; published June 23, 2017 ISSN 1534-7362

The ViCE model did not attempt to address the issue of representation, which was the core of other models (Georgeson, 1994; Georgeson, May, Freeman, & Hesse, 2007; Watt & Morgan, 1983; Watt & Morgan, 1985) . It can be conjectured that there are images in which blur and contrast can be discriminated: in other words, that there are blurred images that have no metamers in the contrast domain. Part of the purpose of the present investigation is to discover the circumstances in which such discrimination is possible, and to see whether it supports the notion of independent "labeled lines" rather than a single channel. The method was to present the observer with a reference stimulus and a single test stimulus (method of single stimuli, MSS) and to require them to press one of four response buttons, to indicate whether the test was of higher/lower sharpness/contrast than the reference. The method is similar to the 2×2FC (forced choice) method of Watson and Robson (1981) and has the same purpose: to provide evidence for labeled lines near to detection threshold. However, the methods are subtly different. Watson and Robson's method is $2 \times$ 2AFC (2-alternative forced choice); ours is MSS with four responses. Watson and Robson's stimuli differed along a single dimension, such as spatial frequency. Ours differed putatively in their sensory dimension, blur versus contrast, although this is moot.

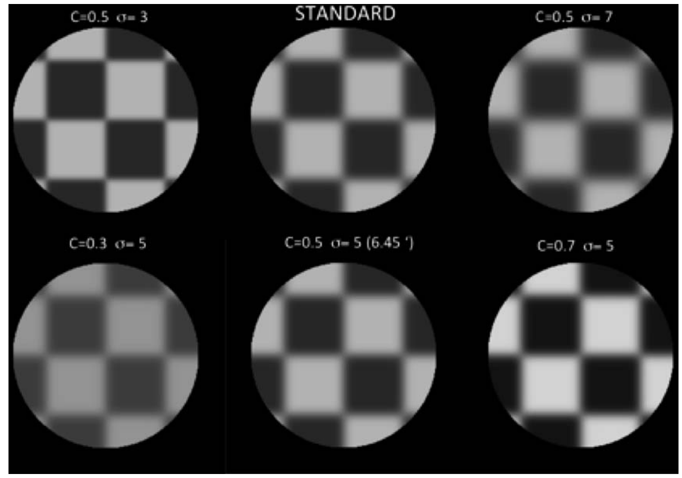

Figure 1. Examples of stimuli used in the experiments. The top row illustrates the standard in the center and two different test blurs (values in pixels; 1 pixel=1.29 arcmin). The bottom row illustrates the standard in the center and two different test contrasts, with the same blur in all cases. The circular aperture is the same as in the actual experiments. The standard had the same contrast (0.50) and blur (5 arcmin) in both contrast discrimination and blur discrimination experiments.

Experiment 1

The purpose of Experiment 1 was to compare blur and contrast discrimination thresholds in a group of observers ($n = 5$) using the same stimuli. The stimuli (illustrated in Figure 1) were chessboards with a baseline contrast of 0.5 and Gaussian blur with a space constant of 5 pixels (6.45 arcmin). Contrast and blur discrimination were measured in separate blocks of trials. On each trial the standard was presented first, followed by the test, and the observer's task was to indicate whether the test had greater or lesser contrast than the standard, or (on blur trials) whether it was more or less sharp. In current signal detection parlance this is the MSS, with a reminder; historically it was called the method of constant stimuli, as used by Weber, although this term has come to take on a different meaning of sampling stimulus levels without replacement.

Apparatus

Stimuli were presented on a Viewsonic PF817 CRT display (Viewsonic Corp., Brea, CA), with pixel resolution 1024×768 and refresh rate 140 Hz and mean luminance 33.5 cd/m^2. Stimulus presentation was controlled by the CRS VISAGE system (Cambridge Research Systems, Rochester, Kent, UK). The display was viewed in a darkened room at approximately 114 cm so that 1 pixel subtended 1.29 arcmin. The stimulus was an 8×8 squares chessboard of alternating white and dark squares with side 1.376° (64 pixels). This was presented within a hard-edged circular mask so that $\sim 3 \times 3$ squares were visible. The purpose of the aperture was to make any edge artifacts invisible when the whole array was filtered. Outside the circle the screen was a maximum luminance red (RGB 1,0,0; $\sim$10 cd/m^2). (This was inherited from a previous eye movement experiment where it was desired to reduce illumination of background features; in retrospect it would have been better to have had a mean luminance surround.) The whole 8×8 squares array of squares was blurred using the MATLAB Image Processing Toolbox "imfilter" routine, with a Gaussian kernel that blurred the edges without affecting the peak contrast:

$$f(x, y) = \exp\left(-\left(x^2 + y^2\right) / \left(2\sigma^2\right)\right) \quad (1)$$

Procedure

The psychophysical procedure was a two-interval MSS with a reference stimulus always presented first before a variable test. The function of the reference was to act as a reminder; note that this was not a 2AFC procedure, where the order of test and reference is randomized. Exposure duration was 1.43 s with a 1.43 s gap in between standard and test during which the display circle was at mean luminance. The reference stimulus for both contrast discrimination and blur discrimination tasks was a chessboard with contrast = 0.5 and blur σ = 6.45 arcmin. (See Fig. 1 for examples). Subjects used the left and right pointing arrow keys on the keyboard to indicate whether the test had less or more contrast than the standard respectively, or (on different blocks of trials) the up and down arrow keys to indicate whether it was "sharper" or "less sharp" than the standard, respectively. Subjects were instructed before each block whether to make the "contrast" or the "sharpness" decision.

In "contrast" blocks of trials (64 trials per block) only the contrast was varied; in "blur" blocks, only the blur varied. In contrast blocks of trials, contrast increment/decrement thresholds in chessboard stimuli (Fig. 1) were measured on a pedestal of 0.5. In blur blocks, blur increment/decrement thresholds were measured on a pedestal of σ = 6.45 arcmin. Feedback was given in the form of a large square in the center of the screen after a correct response and a small square after an incorrect response. The test stimulus on each trial was determined by an adaptive probit estimation procedure (APE; Watt & Andrews, 1981), designed to present stimuli at ±1 standard deviation (s) of the empirical Gaussian psychometric function. For blur thresholds the smallest difference between test and standard was ~0.2 arcmin; for contrast it was ~0.05. Trials with zero difference were also included, and on these trials the feedback was determined randomly. The number of blocks varied between three and eight, according to subject availability.

Subjects

Details of the five observers are given in Table 1. One was the author (MM); the others were unaware of the purpose of the experiment.

N/symbol	Initials	Status	Sex	Age	Naive
1 Red	MM	Author	M	>65	N
2 Green	NN	Postgrad	F	24	N
3 Blue	CC	Student	M	<18	Y
4 Magenta	SA	Student	M	<18	Y
5 Black	ZC	Student	F	<18	Y

Table 1. Details of the subjects used in the Experiments.

Data analysis and modeling

The psychometric functions relating the probability of choosing the test as having higher sharpness/contrast than the standard, as a function of the signed difference between reference and test, were fit with a two-parameter cumulative Gaussian to determine their standard deviation, σ and mean, μ. Examples are seen in Figure 4. The σ value is taken as a measure of the just-noticeable difference, or sensory noise; and μ as any bias the observer has towards seeing the second stimulus as having more/less blur/contrast than the reference.

To transform the data into contrast energy differences in a similar way to the ViCE model, the stimuli used in the experiment were each convolved with the Watson–Ahumada version of the human contrast sensitivity function (HCSF). The contrast energy of each test stimulus, obtained by squaring the HCSF-filtered image and integrating over pixels, was subtracted from that of the standard, to give a signed value D_{ED}, where the subscript indicates the energy–difference model.

$$D_{ED}(t) = \sum_{x=1}^{n}\sum_{y=1}^{n}\left(I(t;x,y) * F(x,y)\right)^2 - \sum_{x=1}^{n}\sum_{y=1}^{n}\left(I(s;x,y) * F(x,y)\right)^2 \quad (2)$$

where $\boldsymbol{I}$ represents an image of n^2 pixels, F is the Watson–Ahumada difference-of-Gaussians filter, t is the test contrast or blur, and s is the standard contrast or blur. The psychometric function was then recalculated using these transformed values instead of the original units of contrast or blur. The standard deviation of the function was taken as the threshold or JND, and finally this JND was transformed into a Weber fraction by dividing it by the standard contrast energy.

It should be noted that this is not exactly the transformation used by Watson and Ahumada. Instead of calculating the energy of the two stimuli separately, and then taking their difference, they measured the energy of the difference between the two filtered stimuli:

$$D_{WA}(t) = \sum_{x=1}^{n}\sum_{y=1}^{n}\left(I(t;x,y) * F(x,y) - I(s;x,y) * F(x,y)\right)^2 \quad (3)$$

We prefer our own calculation in the case of our own stimuli and procedure because the Watson–Ahumada calculation produces an unsigned value for $D_{WA}(t)$ and is thus unable to discriminate whether the test (t) is an increment on the standard or a decrement. However, to be sure that the methods of calculation did not produce different results, we also implemented the Watson–Ahumada calculation, and transformed D into sgn(t).D, where t was the signed test level of contrast or blur.

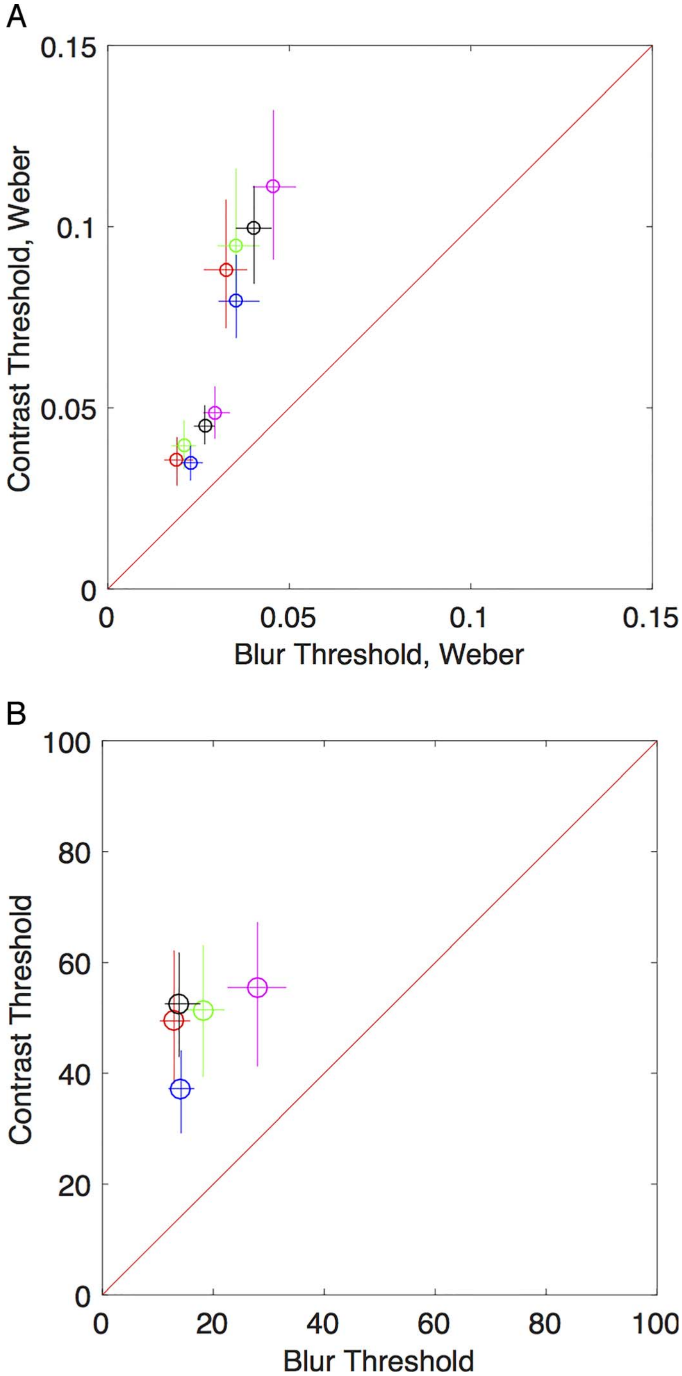

Figure 2. Results of Experiment 1. *Top Panel:* The contrast and blur thresholds were transformed into spatially-filtered contrast energy thresholds between the reference and test using the method described in the text (D_{ED}, Equation 2). These were divided by the reference energy to transform into Weber fractions. Subjects are colored with the scheme in Table 1. The higher cluster shows energy thresholds calculated from the spatial filter used by Watson and Ahumada. The lower cluster values are calculated using a filter with an excitatory space constant that minimizes for each subject the difference between contrast energy and blur energy thresholds. *Bottom panel*: Thresholds were transformed into energies of the →

← difference between the filtered test and threshold stimuli, using the Watson–Ahumada method (D_{WA}, Equation 3). In both panels the error bars are 95% confidence intervals calculated by parametric bootstrapping.

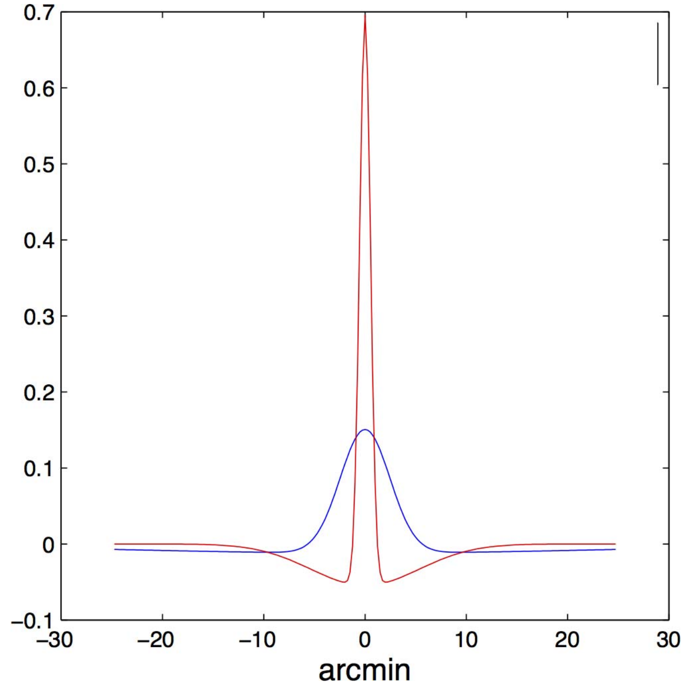

Figure 3. The figure shows the cross-section through the center of the point-spread function for the difference-of-Gaussians (DOG) filter used by Watson and Ahumada to model to human contrast sensitivity function (blue) and a DOG filter (red) used to make contrast and blur discrimination thresholds in Experiment 1 more equal. For further explanation, see the text.

Results

The results (Figure 2) show that blur thresholds were lower than those for contrast, when expressed in equivalent units of visible contrast energy. This confirms the trend that was already apparent in the review by Watson and Ahumada. The figure also shows that they could be made more equal by using a more strongly-band pass filter, the point-spread function of which is shown in Figure 3, in comparison with the almost low-pass filter used by Watson and Ahumada. Even with the fitted filter, however, blur thresholds were significantly lower than those for contrast.

Experiment 2

The purpose of Experiment 2 was to measure the ability of the observers to distinguish changes in blur and contrast. The stimuli were the same as in Experiment 1 but contrast trials were randomly

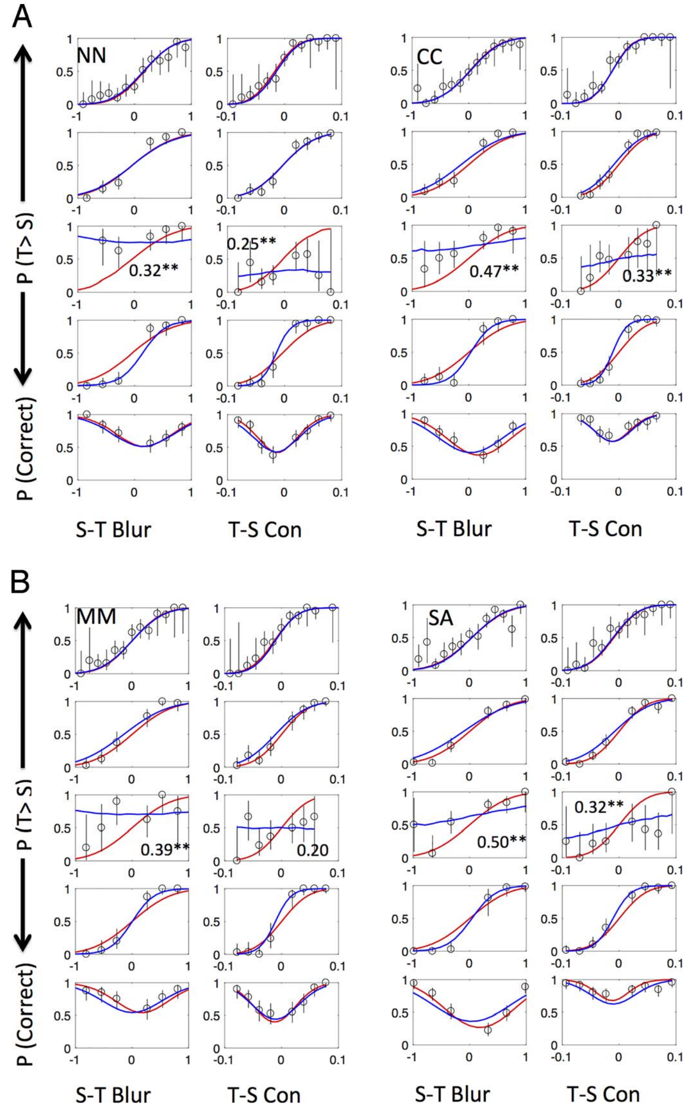

Figure 4. The figure shows the psychometric data from four observers (CC, NN, MM, SA) in Experiments 1 and 2, and the fit of models described in the text. Error bars represent Bayes-credible, 95% binomial confidence intervals (Nicholson, 1985) using code supplied by A. B. Watson. Explanation: The units on the *x* axis are z-values. Each large panel shows the results for a single observer, with initials in the top left hand corner. For each observer the first row shows data from Experiment 1, with blur results on the left and contrast on the right. The ordinate is the probability of choosing the test as having higher sharpness/contrast than the standard. The abscissa is the difference in blur or contrast between standard and test. Note that positive values for blur trials mean that the test was *sharper* than the standard. The second row shows the equivalent data from Experiment 2. The third row shows data from Experiment 2 from trials when the observer made errors in the channel identification task. Row 4 shows equivalent data when the channel identification response was correct. Row 5 shows the probability of a correct channel identification response. Blue curves are the fit of the five-parameter model described in the text, in which sign choice and channel choice depend on the same signals. Red curves are the fits of a nine-parameter model described in the text, in which choice of sign and choice of channel are made from independent signals. Note that both kinds of fit are fits to all the data, not just to the individual psychometric functions in each box.

interleaved with blur-varying trials. Observers used the left and right pointing arrow keys on the keyboard to indicate whether the test was more or less sharp than the standard, and the up and down arrow keys to indicate whether it was of higher or lower contrast. This response mapping meant that both "right" and "up" corresponded with a "more" decision. Only one key press per trial was permitted. Thus, if observers thought that the test was both sharper and of higher contrast than the reference, they had to decide which of the two differences was the greater. Because Experiment 1 had given a good idea of the thresholds for the individual tasks, the adaptive APE procedure was changed to eight fixed stimulus levels, which were sampled randomly without replacement until each had been presented eight times in the course of a session of 128 trials (64 for blur and 64 for contrast). The range of test levels (which did not include zero) was tailored for each observer to span approximately ± 2 JND as determined from Experiment 1. Feedback continued to be used, but a color was added to the square to indicate whether the correct channel (blur or contrast) had been identified. Note that it is possible for the observer to be "correct" for channel and "wrong" for the sign of the change and vice versa.

Subjects, apparatus, and procedure

Except for the differences noted above, Experiments 1 and 2 were the same. Subjects NN, CC, MM, and SA ran seven, seven, five, and nine blocks of 128 trials each, respectively. ZS could not be tested for more than one block, but her data were qualitatively similar to those of the other observers.

Modeling

The model observer has independent channels for blur and contrast, each of which has a signed response. Each channel has additive Gaussian noise, which is scaled by the threshold (σ) in that channel to have unit variance. The signal is scaled by the same amount to become a z-value. The distribution of signals in each channel also has a constant mean offset (μ).

In Experiment 1 (blocked trials of contrast and blur) the observer is assumed to consult only the relevant channel during the block. The psychometric functions relating the probability of choosing the test as having higher sharpness/contrast than the standard, as a function of the signed difference between reference and test, were fit with a two-parameter cumulative Gaussian to determine their deviations, σ_{con}, σ_{blur} and means, μ_{con}, μ_{blur}. However, values for these parameters were fit not only using the data from the blocked trials of Experiment 1 but also using the data from the mixed condition of Experiment 2, in the manner we now describe.

In Experiment 2, the model observer chooses the sign of the signal and chooses between the two channels on the basis of their scalar values on that trial. In the first stage the model observer determines whether the sum of the outputs of the two channels is greater or smaller than zero. (This is equivalent to choosing the sign of the output of the channel with the larger absolute output). Next, the model observer selects which channel to choose (blur vs. contrast) on the basis of the magnitude of the difference in absolute values between the two channels, with an additive response bias. There are five parameters in the model: the variance and mean of the noise in each of two channels, and the bias for channel choice.

The model was used to compute the probability at each signal level of making each of the four possible decisions (+1, +0, −1, −0), where +1 codes for the decision that the signal is positive and in the channel containing the signal and −0 is the decision that sign is negative and the signal is in the wrong channel. These probabilities were determined by a simulation over 10,000 trials. On each trial the scalar value of the channel containing the signal (V1) was sampled from a Gaussian pdf with unit variance and a mean offset of ($\sigma + \mu_1$) where s is the signal value in units of blur or contrast, μ_1 is the mean offset (bias) in that channel also in units of blur or contrast and σ_1 is the standard deviation of noise in that channel. The scalar value of the other channel (V2) is similarly sampled from a Gaussian pdf with unit variance and a mean offset of $(\mu_2)/\sigma_2$. The model observer makes a +ve response if:

(i) $V1 + V2 > 0$

And selects the channel containing the signal if:

(ii) $(\mathrm{abs}(V1)) - \mathrm{abs}(V2) + \mathrm{bias}) > 0$

If both these conditions are satisfied, the decision is coded +1, and so on for the other three possible decisions.

The simulation can easily be implemented in six lines of MATLAB code. An example of the predictions of the model is as follows:

$\sigma = 1$; $\sigma_1 - \sigma_2 = 1$; $\mu_1 = 0.1$; $\mu_2 = 0$; Bias $= 0.2$
p of +1 decision = 0.1318
p of +0 decision = 0.6555
p of -1 decision = 0.0857
p of -0 decision = 0.1270

In a nine-parameter version of the model, the observer has access to two further channels, specified by parameters σ_3, σ_4, μ_3, μ_4 that are used exclusively for making the decision whether the test differs from the standard in contrast or blur. This may seem a bit like buying a dog and doing your own barking, when two independent channels already exist for determining whether the test is higher or lower in contrast/blur than the standard. However, it allows for the possibility that the two kinds of decision are completely independent. A specific prediction of the nine-parameter model is that the probability of correctly deciding higher/lower is the same on trials when the observer is correct/incorrect in choosing contrast versus blur. This is not true in the five-parameter model. The nine-parameter model was easily simulated in exactly the same way as the five-parameter version, except that the two random variables were resampled after one decision and before the other.

On each trial the negative log likelihood of the observer's response was computed from the predicted probability p and the results were added over all n trials:

$$L = \sum_{i=1}^{n} -(\log(p_i)) \quad (4)$$

The quantity L was then minimized with the MATLAB *fminsearch* routine.

An important difference from the model previously described by Raphael and Morgan (2016) for mixed trials of varying texture size and density should be noted. Raphael and Morgan incorporated a response bias for channel choice in their model, but this was done before the decision was made for sign, instead of subsequently as in the present version of the model. The consequence for Raphael and Morgan is that the conditional probability of a positive response given an incorrect choice of channel had to be independent of signal strength. This did not agree with the data for some subjects, who showed clear evidence for sign discrimination even on such trials. In the present version of the model, this is allowed because there will be trials when the channel containing the signal has the correct sign, and has the largest absolute value, but is not chosen in the second (channel choice) stage because of the response bias.

Results

Results for the four main observers in the experiment are shown in Figure 4. Results for the fifth observer

(ZS) , who was unable to complete more than one block of trials in Experiment 2 were qualitatively similar and are shown in the Supplementary Material. The top two rows of the figure show that observers could successfully distinguish increments from decrements relative to the standard both for blur and contrast; and this was true both in the single block conditions (row 1) and when blur and contrast conditions were interleaved (row 2). The bottom row (row 5) shows that observers could also distinguish whether the test differed from the standard in contrast or in blur. Moreover, they could do so about as well as the model predicts from two independent, noisy channels, scaled by the standard deviation of the noise in the channel (blue curve). The bottom row also shows that the success rate was not the same for blur and contrast, because observers had preferences for one or the other. This is accounted for by the bias parameter in the model fit.

Rows 3 and 4 of the Figure show increment/decrement performance on trials when the channel choice (blur vs. contrast) was incorrect (row 3) and correct (row 4). Not surprisingly, performance in the former case is inferior to the second, indicating a common source of noise for the two choices, but it was not at chance in all cases, indicating that observers had some information about the direction of the cue even when they were mistaken about its origin. There was a significant correlation in most cases between the probability of a "sharper" response and an increase in contrast, and between a "high contrast" response and an increase in sharpness. These correlations are shown in the panels of row 3 in Figure 4. A likelihood ratio test (Hoel, Port, & Stone, 1971) comparing a two-parameter linear fit to the data in row 3 with a one-parameter linear fit of unit slope showed a significantly ($p < 0.05$) better fit in all cases except for MM (blur trials) and SA (blur trials).

The mostly positive slopes of the conditional probability data in row 3 of Figure 4 are consistent with observers confusing high contrast with high "sharpness" (low blur). To some extent these positive slopes are explained by the model incorporating channel bias (blue curves), which also tend to have positive slopes. However, it should be emphasized that the sign of the slope is arbitrary as far as the model is concerned, and the slopes would have been negative if we had associated high-contrast responses with high blur, rather than high sharpness. The direction of the effect thus implicates a perceptual effect, consistent with the bias described by May and Georgeson (2007) to see sharper edges as of higher contrast. We shall return to this point in the discussion.

An alternative to the model with two independent channels, accounting for choice of both sign and channel, is one in which choices of sign are made independently of channel. For example, choice of sign could be based on a common contrast energy signal, whereas the choice of channel could depend on a separate mechanism with a higher threshold. This model was tested against the data with a nine-parameter model the same as the 5 parameter except for two further channels used only in the channel-choice stage. These each had independent values of μ and σ from the channels used in the sign discrimination stage, hence the four further parameters. The fits, shown in red in Figure 4 were evidently worse than those of the five-parameter fits, despite the greater number of parameters. The superiority of the five-parameter fit was born out by the X^2 values of a likelihood ratio test [109.37**, 44.99***, 67.02***, 107.74***, and 4.64 (NS)]. Only in observer ZS was the difference not significant, presumably due to the paucity of data in Experiment 2. The failure of the nine-parameter model is explained by its predicting complete independence of the sign and channel choice. Thus the conditional probability of a correct sign choice should be the same given a correct rather than an incorrect channel choice. The data in rows 3 and 4 show that this is clearly not the case.

In most respects, these data are similar to those described by Raphael and Morgan (2016) for discrimination between changes in dot density and dot number. There is a particularly close analogy between the two sets of experiments, since both density and number affect contrast energy, despite which observers can discriminate between them.

Experiment 3

The purpose of Experiment 3 was to investigate summation between contrast and blur cues. Unlike Experiments 1 and 2, a 2AFC procedure was used, with a standard stimulus (blur = 6.45 arcmin; contrast = 0.5) and a test stimulus that had a blur increment ΔB. The order of presentation was randomized and the task was to choose the sharper stimulus (the standard). In the baseline condition the contrast of the standard and test stimuli were the same. In the blur + contrast condition a contrast increment ΔC equal to half the JND (established in Experiment 2) was added to the test. Note that this would have *reduced* the energy difference between test and standard. In the blur–contrast condition a contrast increment ΔC equal to half the JND (established in Experiment 2) was *subtracted* from the test. Note that this would have *increased* the energy difference between test and standard.

Two observers from Experiments 1 and 2 (MM and NN) performed the experiment. The APE procedure was used, as in Experiment 1.

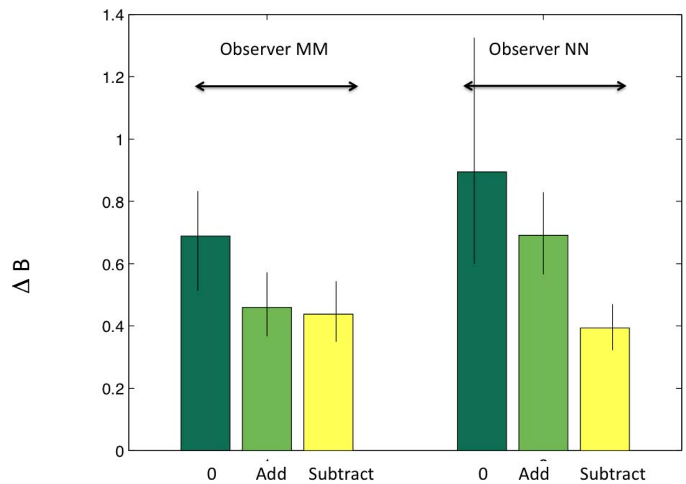

Figure 5. The figure shows the results of Experiment 3 in participants MM (left) and NN (right). The vertical axis shows blur-increment thresholds ΔB. In the baseline condition (0) the contrast of the standard and test stimuli were the same. For explanation of the "Add" and "Subtract" conditions, see the text. The vertical bars show 95% confidence intervals established by bootstrapping.

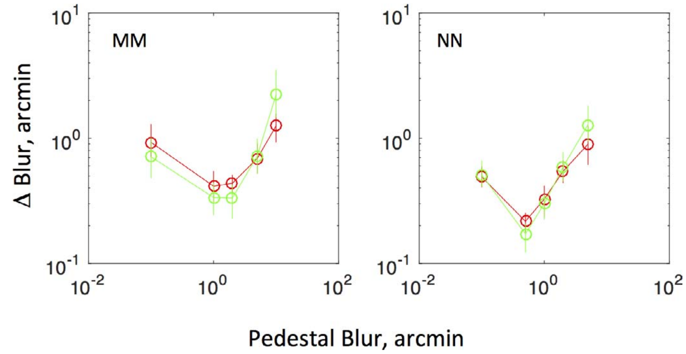

Figure 6. Results of Experiment 4, in which the threshold for detecting an increase in blur in the test was measured in a 2AFC design as a function of the pedestal blur, present in both reference and test stimuli. The red symbols show thresholds when the contrast of test and reference stimuli was the same; the green symbols show thresholds when a contrast increment was added to the test stimuli so as to counteract the decrease in contrast energy due to added blur.

Results

The results in Figure 5 show summation in both observers, in that thresholds were lowered by the contrast cue irrespective of whether it was an increment (+) or a decrement (–). In detail, however, the two observers were different. MM showed significant lowering of threshold relative to baseline in both the add (+) condition $X^2 = 5.22$*) and the subtract (–) condition ($X^2 = 6.24$*) , which did not themselves differ $X^2 = 0.085$ NS). NN showed no significant lowering of threshold relative to baseline in the + condition ($X^2 = 1.61$ NS) but a significant effect in the – condition ($X^2 = 19.63$**); and the + and – conditions were different ($X^2 = 14.76$**). Thus the data for MM give no support for the energy model and are compatible with independent channels for contrast and blur. The data for NN give qualified support for energy, in that blur increments appeared to be reinforced by contrast *decrements*. Even in this observer, contrast increments failed to support the energy prediction that they should increase thresholds for blur increments; in fact, the change was in the opposite direction, albeit nonsignificantly. A possible reason for the difference between observers is that MM knew, but NN did not, that the test could be different from the reference both in blur and in contrast. The instruction was to choose the standard, that is, the sharper stimulus. Seeing a stimulus that was of higher contrast on some trials, NN may have been deceived into thinking that it was also less blurred, and thus made an error. This is pure speculation.

Experiment 4

In the final experiment we measured the full ΔB/B "dipper" function with a 2AFC procedure, both when the test and pedestal stimuli had the same contrast (0.5), and when the test was given a contrast increment above 0.5 to compensate for the decreased contrast energy caused by the increment in blur. The blur discrimination function (ΔB vs. B) in two observers (NN and MM) was determined with added contrast to the test, so that the blur and contrast increments were equal in terms of threshold units (z scores). These z values were determined in a preliminary experiment where (a) the ΔB/B function was measured with a fixed contrast of 0.5 and (b) the ΔC/B function was determined with a fixed pedestal contrast of 0.5. Thus, every decrease in energy due to a blur increase was compensated by an increase in energy due to contrast. The compensation was carried out using z-scores rather than ViCE, because we have already established in Experiment 1 that ViCE is not a good predictor of the relative detectability of blur and contrast.

Results (Figure 6) showed a clear "dipper" function for blur (Hamerly & Dvorak, 1981; Watt & Morgan, 1983; for other references see Watson & Ahumada, 2011) both with and without contrast energy equalization. Thresholds were slightly lower when contrast equalization was added, consistent with the results of the "add" condition of Experiment 3.

General discussion

These experiments were designed to test whether there are separate labeled lines for contrast and blur discrimination. Previous results had told us that both the amount of contrast and the amount of blur in a pattern could be sensed. The question is whether there are two mechanisms or only one. Our results decisively show that there are two mechanisms, particularly the finding that observers can report whether a change in energy is due to a change in contrast or a change in blur. Experiments on the "dipper" (T vs. C) function for blur discrimination have also shown that the dipper survives even when blur has no effect on energy, showing that blur is not just a special case of energy discrimination,

The Watson–Ahumada ViCE model of contrast and blur discrimination is unquestionably an advance on previous models, because it provides a common metric (visible contrast energy) by means of which different tasks can be compared. Before claiming the existence of a special mechanism for a discrimination, simpler possibilities such as energy discrimination should indeed be investigated. For example, we (Morgan & MacLeod, 2014) were inspired by the ViCE model to investigate the case of "numerosity discrimination," and to test the possibility that it is a special case of texture discrimination, based upon energy differences. Morgan, Raphael, Tibber, and Dakin (2014) showed that a mechanism for energy discrimination supplemented with contrast gain control could easily outperform the human observer in a variety of relative numerosity tasks, even with irrelevant differences in blur and contrast between the textures being compared.

However, the ViCE model does not claim that all pattern discriminations are carried out by contrast energy alone. We do not discriminate two faces because they have different contrast energy (Morgan, Ross, & Hayes, 1991; Oppenheim & Lim, 1981), and even if in special cases we did, the discrimination would probably survive random contrast perturbation. This is known to be the case in a variety of simple pattern discrimination tasks (Morgan, 1991; Westheimer, 1979). Morgan and Regan (1987) showed that spatial interval acuity was little affected by random contrast perturbation of one of the two bars in a spatial interval discrimination task. The inference from many experiments of this kind on pattern acuity, with the possible exception of vernier acuity with closely abutting bars (Morgan & Regan, 1987; Parker & Hawken, 1985) is that the critical variable is the distribution of contrast energy over space, not the integrated contrast energy over the pattern.

Blur discrimination is an interesting case, which could be considered a priori either as a pattern acuity task, or as a contrast discrimination. Watt and Morgan (1983) proposed that edge-blur discrimination is a special case of spatial interval discrimination, in which the cue is the spatial separation between stationary points in the second spatial derivative of the blurred edge. The "dipper" function for blur was explained by the existence of "intrinsic blur" from optical and neural blurring. Watson and Ahumada (2011) criticized the notion of "intrinsic blur" in this context, but also point out that it is not very different from the HCSF in their own model. Blur in the Watt and Morgan model is calculated as the spatial variance of the points on the edge compared with a perfectly sharp template. A difference ΔB between them produces a larger difference in total variance between them when it exceeds the intrinsic blur because of the addition of variances under convolution. The same reasoning explains the dipper function found for spatial variance discrimination in regularly spaced dot patterns (Morgan, Chubb, & Solomon, 2008).

The question remains whether blur discrimination has a spatial component or not. Our own results agree with Watson and Ahumada's review in showing that blur discrimination thresholds are lower than those for contrast discrimination when expressed as contrast energy thresholds. This leads to the conclusion that integrated contrast energy cannot be the correct metric for both tasks. A difference in blur between two patterns can be detected without any difference in integrated contrast energy. This is demonstrably the case in Experiment 4, where contrast energy was made invariant with blur by a suitable covariation of contrast, and where a dipper function was still found.

We also found that differences in blur could be discriminated from differences in contrast about as well as would be predicted from two independent labeled lines, having the same intrinsic noise that limits their individual discrimination thresholds (Watson & Robson, 1981). An exception is that when observers made a mistake in identifying the channel (blur vs. contrast) they were above chance at identifying its direction (increment vs. decrement). In other words, increases in sharpness were confused with increases in contrast and vice versa, exactly as predicted by an energy model. However, the energy model predicts that this confusion would be complete, whereas in fact it is rare. A confusion between contrast and blur has also been described by May and Georgeson (2007), who found that sharper edges look higher in contrast. This finding may well account for the finding, in Experiment 2, of above-chance sharpness increment/decrement performance on trials when the channel choice was incorrect: If the test stimulus is sharper than the reference, then it would look higher in contrast, which could lead the subject to sometimes report that it is higher in contrast, rather than higher in sharpness.

May and Georgeson explained their findings by taking a formal model of edge processing (Georgeson et al., 2007), and adding a simple, physiologically plausible, process that could play a role in noise reduction. This could explain why sharper edges look higher in contrast but it does not explain the reciprocal effect we found: that high contrast edges look sharper. May and Georgeson (2007) reported that reducing the contrast caused edges to look sharper (see also Georgeson, 1994).

This finding would actually predict that, as contrast got low, the observer would be more likely to misperceive it as an increase in sharpness, unlike the effect in Experiment 2. However, the effect of contrast on perceived blur reported by May and Georgeson only kicks in substantially at low contrasts and high blurs, so perhaps the contrast in the current experiments was too high and the blur too low for this effect to be seen. In the midrange of contrast that we used ($\sim$0.5) increases in sharpness may cause increases in apparent contrast, and increases in contrast may cause increases in apparent sharpness.

In Experiment 3 we found no clear evidence for a difference in energy-subtract and energy-add modes. This differs from the results reported by Morgan and MacLeod (2014) for summation of numerosity and contrast signals. Second-order sinusoidal gratings composed of visually-resolved dots were modulated either in contrast, in numerosity, or in both. The observer had to distinguish horizontal from vertical modulation. Threshold modulation depths were dramatically higher when contrast and numerosity were modulated in counterphase (from the standpoint of energy) than when they were modulated in phase. The energy model appears to have strong support from these results (see also Morgan et al., 2014), but not from the results reported here for contrast and blur.

We do not deny that contrast energy discrimination may be a rapid and effective heuristic for detecting when the blur of a pattern has changed. Something similar to it works for some kinds of automatic focus in digital cameras, and it may well have evolved in the visual system. Our own method permitted detailed inspection of the pattern edges, and it may have been this that turned the discrimination into a spatial pattern discrimination task. Further work is required to determine the different conditions under which blur discrimination is performed in the pattern discrimination versus the contrast energy domains.

Keywords: contrast, blur, labeled lines

Acknowledgments

Facilities were provided by City, University of London. The paper was supported in part by a grant from the Engineering and Physical Sciences Research Council.

Commercial relationships: none.
Corresponding author: Michael J. Morgan.
Email: Michael.Morgan@sf.mpg.de.
Address: City University London, London, UK; and Max-Planck Institute for Metabolism Research, Cologne, Germany.

References

Georgeson, M. A. (1994). From filters to features: Location, orientation, contrast and blur. In *Higher-order processing in the visual system* (Ciba Foundation Symposium 184; pp. 147–165). Chichester, England: Wiley.

Georgeson M. A., May, K. A., Freeman, T. C., & Hesse, G. S. (2007). From filters to features: Scale-space analysis of edge and blur coding in human vision. *Journal of Vision*, *7*(13):7, 1–21, doi:10.1167/7.13.7. [PubMed] [Article]

Hamerly, J. R., & Dvorak, C. A. (1981). Detection and discrimination of blur in edges and lines. *Journal of the Optical Society of America*, *71*, 448, doi:10.1364/ JOSA.71.000448.

Hoel, P. G., Port, S. C., & Stone, C. J. (1971). *Introduction to statistical theory*. Boston, MA: Houghton Mifflin.

May, K.A. & Georgeson, M.A. (2007). Blurred edges look faint, and faint edges look sharp: The effect of a gradient threshold in a multi-scale edge coding model. *Vision Research*, *47*(13), 1705–1720.

Morgan, M., & Regan, D. (1987). Opponent model for line interval discrimination: interval and vernier performance compared. *Vision Research*, *27*, 107–118.

Morgan, M. J. (1991). Hyperacuity. In D. Regan (Ed.), *Spatial vision* (pp. 87–110). London, England: MacMillan.

Morgan, M. J., Chubb, C., & Solomon, J. A. (2008). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11):9, 1–8, doi:10.1167/8.11.9. [PubMed] [Article]

Morgan, M. J., & MacLeod, D. I. A. (2014). Non-orthogonal channels for relative numerosity and contrast detection. *Journal of Vision*, *14*(10):392, doi:10.1167/14.10.392. [Abstract]

Morgan, M. J., Ross, J., & Hayes, A. (1991). The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biological Cybernetics*, *65*, 113–119.

Morgan, M. J., Raphael, S., Tibber, M., & Dakin, S.C. (2014). A texture-processing model of the 'visual sense of number.' *Proceedings of the Royal Society of London B*: *Biological Sciences*, *281*, 1790.

Nicholson, B. J. (1985). On the F-distribution for calculating Bayes-credible intervals for fraction nonconforming. *IEEE Transactions on Reliability*, *R34*(3), 227–228.

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, *69*, 529–541.

Parker, A., & Hawken, M. (1985). Capabilities of monkey cortical cells in spatial-resolution tasks. *Journal of the Optical Society of America A*, *2*(7): 1101–1114.

Raphael, S., & Morgan, M. J. (2016). The computation of relative numerosity, size and density. *Vision Research*, *124*, 15–23.

Watson, A. B., & Ahumada, A. J. (2011). Blur clarified: A review and synthesis of blur discrimination. *Journal of Vision*, *11*(5):10, 1–23, doi:10.1167/11.5.10. [PubMed] [Article]

Watson, A. B., Ahumada, A. J., & Farrell, J. (1986). Window of visibility: A psychophysical theory of fidelity in time-sampled motion displays. *Journal of the Optical Society of America A*, *3*, 300–307.

Watson, A. B., & Robson, J. G. (1981) Discrimination at threshold: Labeled detectors in human vision. *Vision Research*, *21*(7), 1115–1122.

Watt, R. J., & Andrews, D. (1981). APE: Adaptive probit estimation of psychometric functions. *Current Psychological Reviews*, *1*(2), 205–213.

Watt, R. J., & Morgan, M. J. (1983). The recognition and representation of edge blur: Evidence for spatial primitives in vision. *Vision Research*, *23*, 1465–1477.

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research*, *25*, 1661–1674.

Westheimer, G. G. (1979). The spatial sense of the eye. *Investigative Ophthalmology & Visual Science*, *18*, 893–912. [PubMed][Article]