Perceived contrast in complex images

Andrew M. Haun
Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA

Eli Peli
Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA

To understand how different spatial frequencies contribute to the overall perceived contrast of complex, broadband photographic images, we adapted the classification image paradigm. Using natural images as stimuli, we randomly varied relative contrast amplitude at different spatial frequencies and had human subjects determine which images had higher contrast. Then, we determined how the random variations corresponded with the human judgments. We found that the overall contrast of an image is disproportionately determined by how much contrast is between 1 and 6 c/°, around the peak of the contrast sensitivity function (CSF). We then employed the basic components of contrast psychophysics modeling to show that the CSF alone is not enough to account for our results and that an increase in gain control strength toward low spatial frequencies is necessary. One important consequence of this is that *contrast constancy*, the apparent independence of suprathreshold perceived contrast and spatial frequency, will not hold during viewing of natural images. We also found that images with darker low-luminance regions tended to be judged as having higher overall contrast, which we interpret as the consequence of darker local backgrounds resulting in higher band-limited contrast response in the visual system.

Introduction

In seeing a scene, we are sensing and binding together image features into objects that inform us about the physical state of the world—this is presumably the purpose of vision. In addition to their utility, these features and objects of vision have a phenomenal impact or magnitude—their perceived brightness or contrast—that correlates with the physical intensity or gradient of the proximal stimulus (the stimulus strength). These perceived magnitudes seem to be determined largely by the relative magnitude of response of the neurons encoding the object or feature; e.g., the perceived contrast magnitude of a spatial pattern seems to closely correspond with the magnitude of neural response in the primary visual cortex (Boynton, Demb, Glover, & Heeger, 1999; Campbell & Kulikowski, 1972; Haynes, Roth, Stadler, & Heinze, 2003; Kwon, Legge, Fang, Cheong, & He, 2009; Ross & Speed, 1991). These responses are not fixed functions of stimulus strength but are subject to numerous nonlinearities, many of which are dependent on spatiotemporal context.

In the phenomenon known as *contrast constancy*, for contrast strength that sufficiently exceeds visual detection thresholds, perceived contrast magnitude is independent of spatial frequency (Brady & Field, 1995; Cannon, 1985; Georgeson & Sullivan, 1975). This behavior is consistent with measurements of contrast discrimination that suggest that the subjective "decision variable" dependent on contrast converges across spatial frequency at high contrasts (Bradley & Ohzawa, 1986; Swanson, Wilson, & Giese, 1984). Contrast constancy has been demonstrated through the use of narrowband stimuli like gratings or band-pass noise with the intention that individual mechanisms in the visual system may be studied in isolation and their properties compared with one another as independent perceptual devices. However, it has been clear for many years that these mechanisms are not independent. Different simultaneously activated mechanisms suppress one another, with these suppressive processes usually held up as types of response normalization (Blakeslee & McCourt, 2004; Foley, 1994; Graham, 2011; Graham & Sutter, 2000; Watson & Solomon, 1997; Wilson & Humanski, 1993). Because natural scenes consist of many simultaneous, overlapping stimuli, perception of natural scenes must be rife with suppressive interactions. Threshold measurements made against broadband scene or noise backgrounds (or immediately after adaptation to these) suggest that perceptual responses to lower spatial frequency contrasts are disproportionately suppressed relative to higher spatial frequencies (Bex, Solomon, & Dakin, 2009; Haun

Citation: Haun, A. M., & Peli, E. (2013). Perceived contrast in complex images. *Journal of Vision*, *13*(13):3, 1–21, http://www.journalofvision.org/content/13/13/3, doi:10.1167/13.13.3.

doi: 10.1167/13.13.3 Received May 8, 2013; published November 4, 2013 ISSN 1534-7362

& Essock, 2010; Webster & Miyahara, 1997) although this has also been interpreted as increasing susceptibility to noise at low frequencies (Schofield & Georgeson, 2003). These findings call into question the matter of contrast constancy in naturalistic stimulus contexts.

The difficulty in determining whether contrast constancy, or something else, occurs in scene perception is that broadband imagery does not appear, to the observer, as a set of identifiable mechanism responses. Rather, the broadband percept is unified across spatial frequency: Virtually all models of broadband feature perception involve collapsing transduced contrasts across spatial frequency before perceptual judgments are made (Georgeson, May, Freeman, & Hesse, 2007; Kingdom & Moulden, 1992; Marr & Hildreth, 1980; Peli, 2002; Watt & Morgan, 1985) so that broadband features—edges and textures—are seen holistically. This holistic percept is compulsory, and no amount of effort will allow an observer to "see through" an edge so that its components can be perceived separately. This phenomenal opacity of broadband features means that any judgment of the perceived contrast of a component frequency would be confounded with judgments of other spatial frequency contrasts. Testing sensitivity (signal-to-noise ratio) to each component within a broadband structure, which has been done in numerous contexts (Bex et al., 2009; Haun & Essock, 2010; Huang, Maehara, May, & Hess, 2012; Schofield & Georgeson, 2003), fails to disambiguate the contributions of internal noise and response magnitude, and the response magnitude is exactly what we want to discover. Our solution to this problem—how to measure the perceived contrast of the spatial frequency *components* of complex images—is to have observers judge the contrast of an entire broadband image without the requirement that particular attention be paid to one or another spatial frequency. To discover how the different components of broadband images contribute to their perceived contrast, we adapted the classification image paradigm (Ahumada & Lovell, 1971; Beard & Ahumada, 1998), in which the stimulus is randomly varied over many trials while observers make simple judgments requiring consideration of the information being varied, and correlations are sought between the stimulus variation and the observer's decisions. Such reverse correlation methods have been used to reveal spatial frequency tuning functions for detection of white noise (Levi, Klein, & Chen, 2005) and to better understand detection of narrowband signals *in* visual noise (Taylor, Bennett, & Sekuler, 2009). Here, we make use of a similar technique to derive spatial frequency weighting functions for subjective estimates of broadband image contrast.

In our experiment, observers are presented with two copies of an image, each of which has had its amplitude spectrum independently modified, and are asked to choose which image seems to be of higher contrast. Over many trials, this procedure provides a probe of the observer's own perceived contrast-weighting scheme. We find that when viewing natural scenes, subjects behave as though perceived contrast is not constant with spatial frequency. Rather, subjects behave as though spatial frequencies around the peak of the CSF are seen as contributing most to the overall impression of image contrast, and low and high spatial frequencies contribute less. We find that the perceptual contrast-weighting scheme is largely independent of image structure, suggesting that the form of the broadband impact of natural imagery is a robust feature of visual perception. Through simulations, we test several models of perceived contrast and conclude that a standard multichannel system with a contrast-gain control that strengthens toward low spatial frequencies is consistent with our data.

Methods

Subjects

Six subjects, aged 22 to 32, participated after giving informed consent in accordance with the Schepens IRB requirements and the Declaration of Helsinki. Subject AH was the first author; all other subjects were naive as to the purpose of the experiment. All subjects had normal or corrected-to-normal visual acuity and no other known visual problems. The experiment was carried out in a darkened room. The display was viewed binocularly.

Stimuli

Stimuli were 516 digital photographs of scenes, including indoor and outdoor, natural and man-made content (Hansen & Essock, 2004). Images were resized from 1024 × 1024 to 768 × 768 pixels to exclude noise at the highest spatial frequencies and then cropped to allow the whole image to fit on the display (as shown in Figure 1; the area of the mirror images inside and outside the green rectangle on either side of the central divider is 768 × 504 pixels). In each trial of the experiment, an image was selected, cropped to 480 × 480 pixels, and divided into eight bands, including seven one-octave bands defined by a raised cosine of log frequency envelope h (Equation 1) with peak spatial frequencies for the first seven bands at 1, 2, 4, 8, 16, 32, and 64 cycles per picture (cpp). Cosine filters were used because they can be fit back together without distortion of the original image (cf. Peli, 1990).

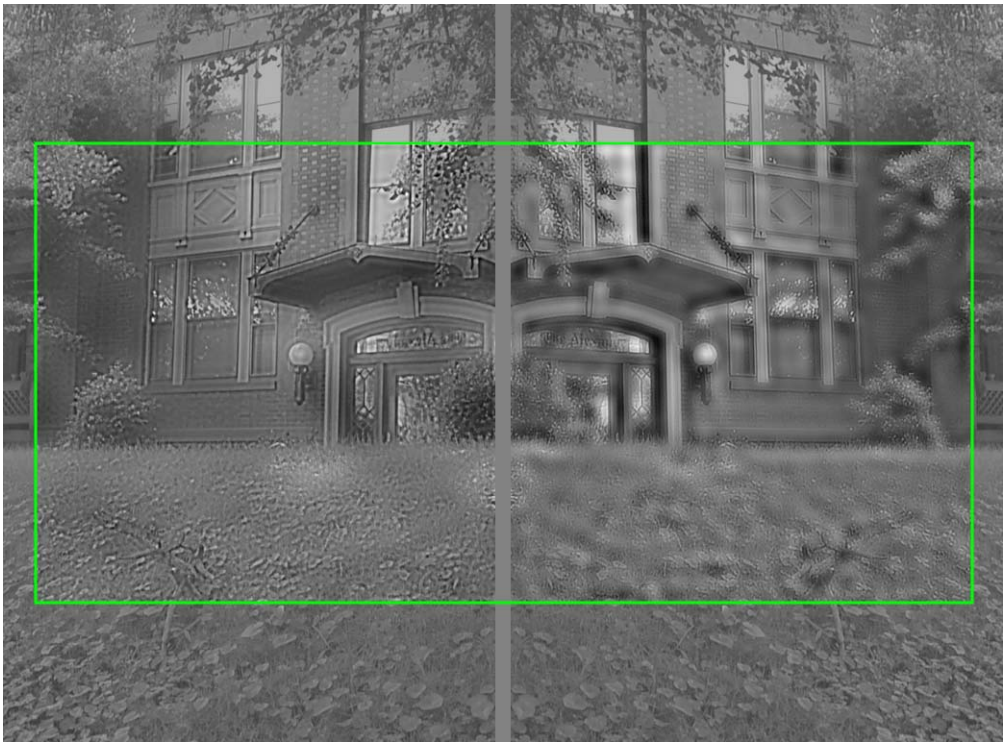

Figure 1. Experiment display. Total display area was 16.5° × 22°. The test areas were 10.3° × 10.3°, presented on either side of a 16-pixel gray divider as mirror images. The test area here is indicated by a green bounding rectangle, which was presented for 1 s at the beginning of each trial to remind subjects of the extent of the test area. The remaining display area was filled with unaltered surround structure from the source images. The test image shown here corresponds to the first set of weights (T1) in Figure 2.

$$h(s) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\pi\log_2(f) - \pi\log_2(s)\right)\right], \\ \qquad if\left\{\frac{1}{2}s < f < 2s\right\} \\ 0, \qquad otherwise. \end{cases} \quad (1)$$

Here, f is spatial frequency, and s is the center spatial frequency of the filter. At the viewing distance of 1 m, a 480-px image subtended 10.3° of visual angle, so the spatial frequency bands were centered at 0.1, 0.19, 0.39, 0.78, 1.6, 3.1, and 6.2 cycles per degree (cpd). The remaining high spatial–frequency content was assigned to an eighth "residual" band, which plateaued at 128 cpp or 12.4 cpd; i.e., for $f \geq 12.4$, $h(s) = 1$. Two vectors of random reweighting coefficients were generated ranging from −8 dB to +8 dB (a decibel is 20 times log10 [contrast amplitude]) and applied in order to the series of frequency bands to produce two reweighted series, which were summed to obtain two altered versions of the original image. The two images were jointly normalized to fit within the 0–1 display range, and the pair was displayed as seen in Figure 1 and as described below in the Procedure section. To produce the random coefficients, the following algorithm was followed:

1. Create a vector $\boldsymbol{\omega 1}$ of eight uniformly distributed random coefficients, normalized so that the maximum value is equal to +8 dB and the minimum value is equal to −8 dB.
2. Create a second vector $\boldsymbol{\omega 2}$ by randomly rearranging the order of the values in $\boldsymbol{\omega 1}$.
3. If the positive or negative maximum coefficients have the same location in $\boldsymbol{\omega 1}$ and $\boldsymbol{\omega 2}$, repeat step 2.

Thus the two copies of the stimulus scene had the same relative amount of contrast change (absolute contrast change would depend on the specific structure of the scene; i.e., the two resulting images did not have the same RMS contrast), and the largest changes were never the same in both copies (e.g., the same band could not be increased by 8 dB in both copies, but it could be increased by 8 dB in one and *decreased* by 8 dB in the other). The range of ±8 dB was chosen because it was large enough to allow contrast differences to be easily visible between stimulus pairs (narrowband contrast discrimination thresholds tend to be about 1 or 2 dB relative to background contrast) while not completely disrupting the appearance of the scenes: While the band-randomized images are distorted, their content is clearly recognizable (Figure 1). A schematic illustration of sample band-weight pairs, over three consecutive trials, is shown in Figure 2.

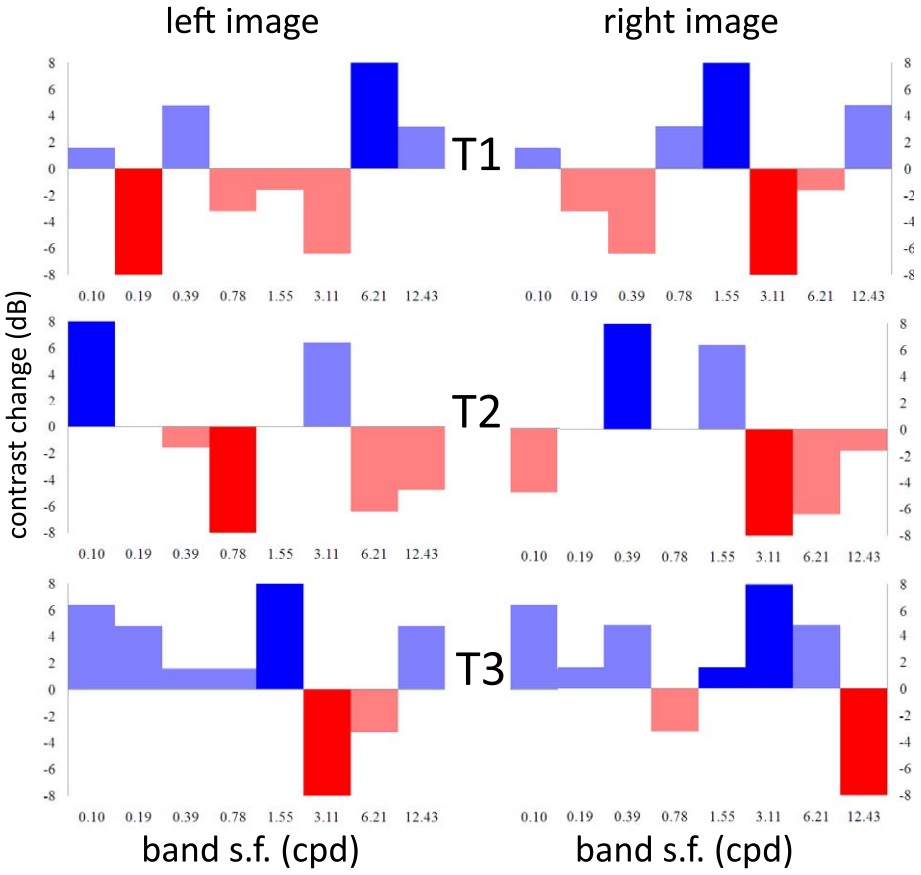

Figure 2. Contrast weights generated for three experiment trials Ti, illustrating the test image comparison as between the contrast-randomization vectors. Within each trial (each row), the same changes in contrast are applied to each test image but in different order. Blue bars are increases in band contrast; red bars are decreases. The longest bars represent the maximum deviations of 8 dB.

The remaining image area (outside the green bounding rectangle in Figure 1) was used to provide a textured, contiguous surround to the test stimuli (except for the central 16-pixel strip, which was featureless, to make clear the separation between left- and right-side stimuli). The contiguous surround was used to avoid having to treat the boundaries of the test images as edges, preventing, e.g., a sudden drop in surround suppression at the boundaries. The cropped area was from the inner sides of the image pairs as shown in Figure 1. The mean luminance of the display was allowed to vary from trial to trial although each pair of test images was constrained to have the same mean. This was done to make maximum use of the display bit depth.

Procedure

In each trial, the two copies of the source scene were presented as mirror images with one of the two (randomly selected) flipped from left to right (as in Figure 1). A thin, 4-pixel mean luminance frame separated the test images from their contiguous surrounds and was highlighted in green for the first second of each trial. Trials were untimed, and subjects were instructed to explore both images in order to decide which one had higher contrast (i.e., eye movements were allowed). "Contrast" was explained as "the range of grayscale values you see in the image; brighter bright areas and darker dark areas indicate higher contrast." Subjects were instructed to take the entire area of the test images into account in making their judgment. The distinction between contrast and sharpness was pointed out to each subject; i.e., that blurry images could potentially have a larger range of brightnesses than sharp images, but the distinction was not stressed because we did not want to induce selection *against* sharper images. To confirm that the results presented below were not due to this instruction, two more subjects were run through the experiment and given only the instruction to choose the image with higher contrast without making the blur/contrast distinction. Subjects were given the basic task of

choosing left or right—which image had higher contrast—but for each choice, they were required to also determine whether their choice was obvious or difficult so that their final response took the form of pressing one of four keys: strong left, weak left, weak right, strong right. The main experiment was run in four blocks of 500 trials each, drawing stimuli from the set of 516 scenes without replacement within each block.

Equipment

The experiment was implemented using Matlab 7.5 with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The display was a linearized (through the video-card software) Trinitron CRT with mean luminance 39 cd/m^2, run at 768 × 1024 resolution (2.67 px/mm) and a 100-Hz frame rate. For the threshold measurements (Appendix B), color bit–stealing (Tyler, 1997) was used to obtain ~10.4 bits of grayscale resolution. In the main experiment, grayscale resolution was 8 bits (bit-stealing was not used). The monitor settings (RGB luminance responses) were the same in all conditions.

Modeling and simulation

There are standard models of contrast perception that account for numerous facts relating to performance measures, like detection and discrimination thresholds, and subjective measures of perceived contrast. Before evaluating the results of the experiment, we will here introduce a series of these models whose performance can be compared with the human subjects. The models were devised as described in detail in Appendix A and then run through the same experiment as the human subjects. To ensure that the model-human comparisons were based on similar image inputs and similar perceptual sensitivities, we used the average of the human subjects' CSFs to calibrate the models' observers, and the modulation transfer function (MTF) of our display was applied to the model stimulus images (procedures for these measurements are described in Appendix B). Each model was run through 4,000 trials of the experiment.

RMS observer

Our human subjects were instructed to estimate the range of grayscale values in the test images; in essence, they were being asked to estimate the RMS contrast of the images. So an observer with no biases or limitations should just compute the RMS contrast of the two test images in each trial and choose the image with the larger value. This simple model doesn't behave like a human observer at all. It is referred to below as the *RMSob*.

CSF observer

A better bet than the RMSob is a model of human contrast perception, which we will refer to as the *CSFob*. The CSFob incorporates a contrast threshold function (the CSF) that sets the minimum threshold, as a function of spatial frequency, for stimulation of the system: an array of independent spatial frequency–tuned channels or filters and a compressive supra-threshold transducer function. Here filter amplitude (in the frequency domain) is constant with frequency, corresponding to *spatial* filter amplitude or sensitivity that increases with frequency (Field, 1987; Georgeson & Sullivan, 1975; Kingdom & Moulden, 1992). Importantly, this model predicts contrast constancy for high-contrast, narrowband patterns of different spatial frequency as well as for 1/f broadband images (Brady & Field, 1995). We have generally adopted the formulation of Cannon's perceived contrast model (Cannon, 1995; Cannon & Fullenkamp, 1991) as the basis for our CSFob. The structure of this model is illustrated in Figure 3. The CSFob doesn't reproduce human behavior very closely, but it is the basis for the following elaborations:

CSF + "white" gain control

Depending on the source (from surround or overlay masks, across frequency or orientation, etc.), masking takes different forms, but the most general seems to be the contrast gain–control model described by Foley (1994) and developed in a similar form by many others. Foley's model is expressed in a simple function that can be equated with sensitivity (d'):

$$d' = r\frac{C^{p+q}}{z^p + \sum_i w_i C_i^p}. \quad (2)$$

Here, C is the linear response of a band-pass spatial filter, and the other parameters are fixed constants (linked to the filter in the numerator). The denominator describes a summation of inputs to the mechanism's gain control, including a constant term z that sets minimum contrast sensitivity, and inputs from nearby mechanisms i. This form of extrinsic gain control shifts the low-contrast part of the transducer function toward higher contrasts, but the transducer still converges to similar levels for high contrasts, so, for high-contrast stimuli, contrast constancy can still be obtained if the

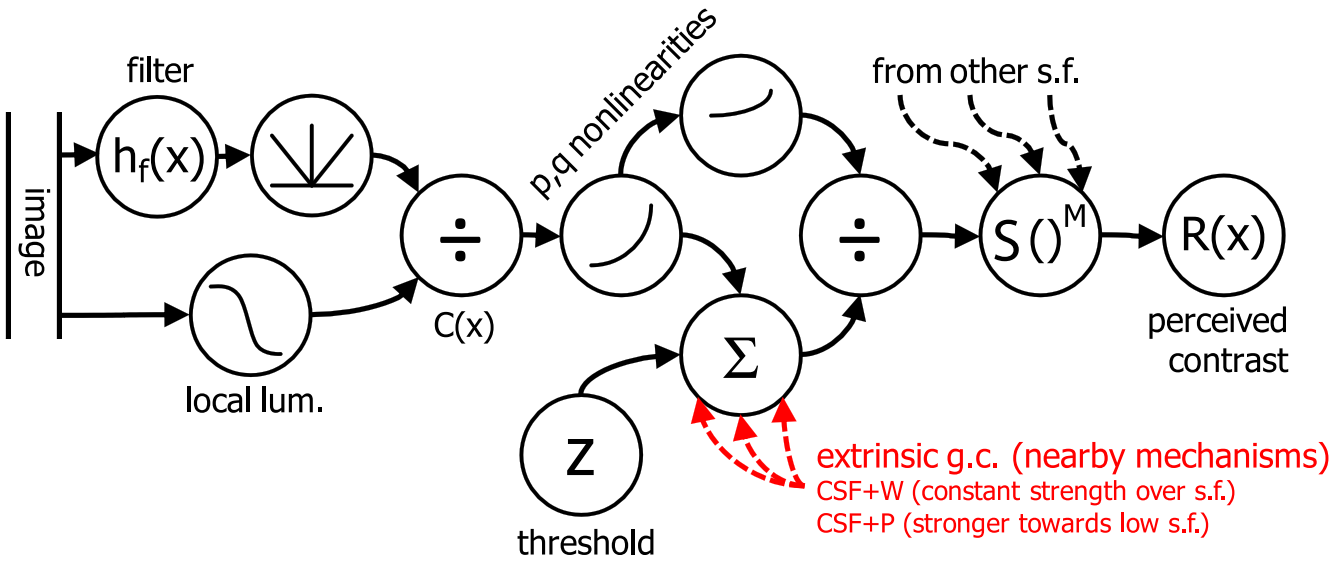

Figure 3. The model of contrast perception used in our simulations at a particular image location *x* and specific spatial frequency (*f*). The nodes represent operations in the order (from left to right) implied by the model described in Appendix A (especially by Equations 2 and A4), beginning with a linear measure of stimulus contrast and adjustment by the local luminance, followed by expansive nonlinearities and divisive gain control fed by intrinsic and extrinsic factors and ending with a summation of responses from other mechanisms over spatial frequency. At this stage, perceived contrast judgments *R(x)* can be made for arbitrary spatial locations. The crucial stage in explaining our experimental results is highlighted in red: The weighting of the extrinsic gain control over spatial frequency can be set according to different rules. Our results suggest that the gain control is stronger for mechanisms that transduce lower spatial frequencies (CSF + P).

maskers are not too powerful. These models are often found in the front ends of image-quality metrics or other applied vision algorithms (beginning with Teo & Heeger, 1994; reviewed in Haun & Peli, 2013b). It's still unclear just how this gain control should be weighted for different mechanisms—i.e., how w_i should depend on stimulus frequency or orientation—but a good null hypothesis (which has generally gone unstated) might be that it is the same everywhere (Teo & Heeger, 1994; Watson & Solomon, 1997). Because the gain control is flat over all stimulus frequencies, we refer to it as *white* gain control, so this model is the $CSF + W$.

CSF + "pink" gain control

The assumption of flat gain-control weights over all stimulus dimensions is probably wrong. Some recent studies have proposed that there is a low spatial–frequency bias in the strength of contrast-gain control (Hansen & Hess, 2012; Haun & Essock, 2010) or a high-speed bias (Meese & Holmes, 2007); some have suggested that gain-control weights are also anisotropic with orientation (Essock, Haun, & Kim, 2009; Hansen, Essock, Zheng, & DeFord, 2003; Haun & Essock, 2010). To incorporate a low spatial–frequency bias into the masking model (our experiment design provided no information about orientation), we weighted the gain control with a negative power function of spatial frequency (Meese & Holmes, 2007). This model, which reproduces human performance very closely in most respects, we refer to as the *pink* gain-control model or $CSF + P$.

CSF + band-limited contrast

The computation of spatially local contrast in an image should be done with respect to the local mean luminance (Kingdom & Moulden, 1992; Peli, 1990) rather than relative to the global mean. To reproduce all of the features of human performance in our experiment, we found it necessary to include such a band-limited contrast computation. Except in one case specified below, simulations were performed with band-limited contrast inputs.

Results 1: Decision weighting functions

Decision weights

To understand how subjects were using the contrast in different frequency bands to make their decisions about overall scene contrast, we compared the band

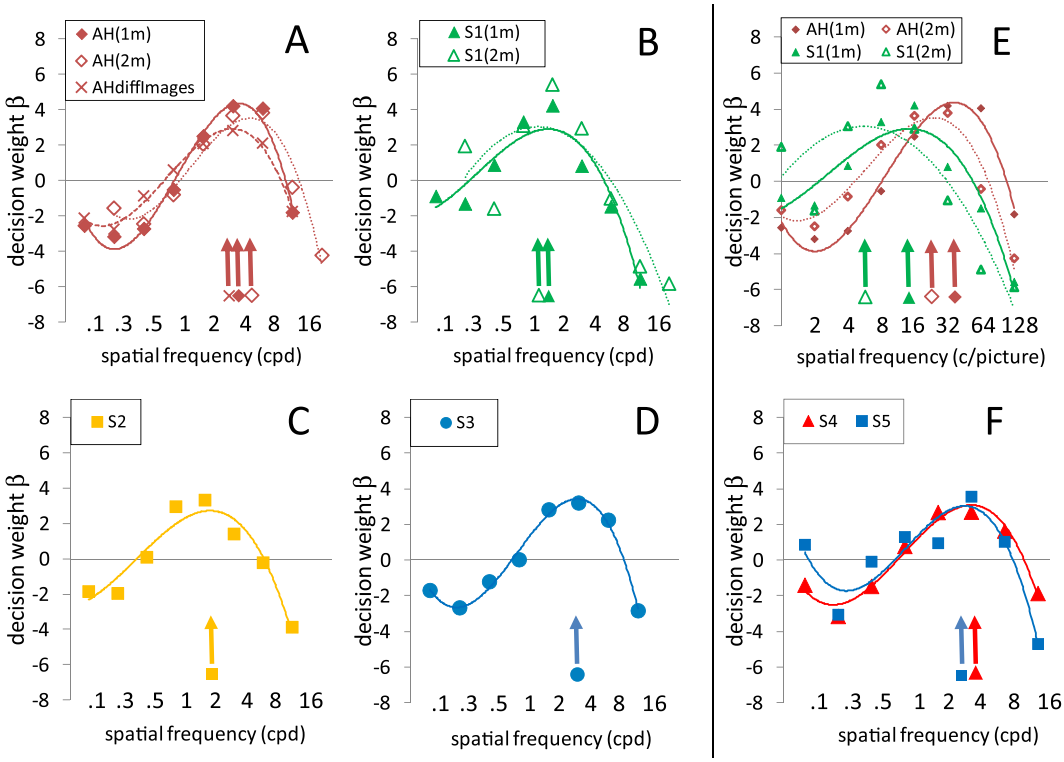

Figure 4. (a–d) The y-axis is the coefficient β from Equation 3 that relates subject choices as to which image has higher contrast to fluctuations in band contrast. The x-axis plots spatial frequency in cpd except for in (e). Solid symbols are data from the main experiment. Open symbols in (a) and (b) are for the 2 m viewing condition. The "x" symbols in (a) are for a control condition in which the two test images in each trial were not the same. The curves are third-degree polynomials, used for interpolating the peaks of the weighting functions. The arrows near the x-axes indicate the "peak weighted frequency," the argmax of the curves. (e) Replots the 1 m and 2 m data from (a) and (b) as a function of cpp. The evident shift in weighting function position between viewing distances shows that it is retinal spatial frequency that matters to the subjects. (f) Data for two subjects who were not informed as to the separability of image blur and contrast.

contrast in images that were chosen versus images that were not chosen, weighted by the subjects' confidence in their decisions. We did this by looking at the average ratio of the contrasts in chosen versus rejected stimuli (the difference in decibels as shown in Equation 3). Because the two test images had the same native frequency spectra (before the experimental manipulation), the base contrasts divide out, and we only need compare the multiplicative weights ω, over all trials T:

$$\beta_f = \frac{1}{T}\sum_{i=1}^{T}\gamma(\omega_{f,i}^{choose} - \omega_{f,i}^{reject}). \quad (3)$$

We treated each subject's strong/weak judgments as equally spaced four-point ratings of each image in each pair, so that a strongly chosen image was rated as 4 and necessarily paired with a strongly rejected image rated as 1, with weakly chosen/rejected image pairs taking the intermediate values (3 and 2). The value γ represented these ratings as a weight in the summation of Equation 3 and was set to 3 (4 minus 1) for strongly discriminated image pairs and set to 1 (3 minus 2) for weakly discriminated image pairs. We should stress that the technique described by Equation 3 is simply a means of describing subject performance in the task and *not* a model of contrast perception. Positive and negative values of β describe how contrast at each band influenced the subjects' *decisions about* perceived image contrast and do not describe perceived contrast directly.

Decision weights β_f are plotted in Figure 4a through d for each subject. Each function peaked at spatial frequencies between 1.5 and 6.0 cpd. Two of the subjects (S1 and S3) had weighting functions peaking closer to 1.5 cpd with the other two peaking between 3.0 and 6.0 cpd (author AH and S2). Coefficients for all subjects were negative for content in the lowest two spatial frequency bands (.09 and 0.19 cpd) and for the high-frequency residual band (12.0 + cpd). All subjects had positive coefficients for the 1.5-cpd band. The main experiment was conducted at a distance of 1 m, but 1,000 trials were also collected for two subjects (S1 and AH) at 2 m with no changes to the display. The 1 and 2 m data are replotted in Figure 4e as functions of image

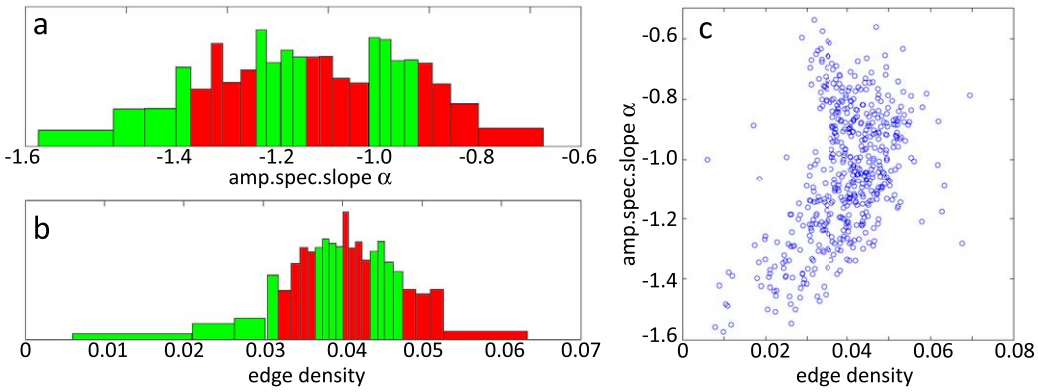

Figure 5. Basic image statistics for the (unaltered 480 × 480 source) test images. (a) Log-log slope α of the radially averaged amplitude spectrum. The area of each bar represents a similar proportion of the test images (21 or 22 of the 516 images in alternation) covering the range indicated on the abscissa. The alternating colors demarcate the six bins for which separate weighting functions were calculated in Figure 6. (b) Spatial average of the Canny-filtered images, interpreted here as "edge density" with the same convention as in (a). (c) Scatter plot of the 516 test images showing that there is only a moderate correlation (= .447) between the two statistics.

frequency (cpp). The functions appear aligned to angular retinal frequency (cpd) (in Figure 4a and b) rather than image frequency (cpp). From this, we conclude that the proper axis on which to present our results is in units of cycles per degree of visual angle. Figure 4f shows data for two additional subjects who were not instructed as to the distinction between contrast and blur; the weighting functions for these two subjects closely resemble those of the original four subjects. Figure 4a also includes data from a version of the experiment in which the two test images were not drawn from the same source image ("x" symbols); this weighting function resembles the one from the main experiment although its amplitude is less (because "strong" choices were less frequent: this version of the experiment is much more difficult).

Results 2: Importance of scene statistics

By measuring the weighting coefficients over all trials of the experiment, we are assuming that only the random weighting vectors (ω) were varied. However, the scene content was not controlled, and a wide variation in content was present by displaying different scenes in each trial and by allowing subjects to freely explore the stimuli. Complex image structure interacts with contrast perception through phase structure (Huang et al., 2012), overlay and surround masking (Chubb, Sperling, & Solomon, 1989; Kim, Haun, & Essock, 2010), and orientation biases (Essock, DeFord, Hansen, & Sinai, 2003). Different images will contain different textures and different edges and surfaces at different scales and orientations, so it is reasonable to suppose that scene structure may influence global estimates of image contrast. The large spatial extent of the test images, the instructions given to the subjects that they should make every effort to consider the entire area of the test stimuli, and the likely large individual variation in how they satisfied this instruction, make consideration of local scene statistics untenable. Instead, we consider here the effects of global scene statistics on the weighting of different spatial scales in judgments of perceived contrast.

Amplitude spectrum slope

The slope α of the radially averaged spatial frequency amplitude spectrum is a useful scalar measure of the general appearance of an image. This statistic describes the distribution of physical contrast over spatial frequency, thus predicting the spatial frequency distribution of contrast responses of an image-sensing system (Field, 1987). Measured in the Fourier domain, α is the exponent of a power function fitted to the radially averaged amplitude spectrum. We did this for each (undistorted source) test image (Figure 5a). For each subject, trials were sorted into six bins (333 or 334 trials each) according to the α value of the original test image for each trial. We measured decision-weighting functions for each bin using Equation 3, and the peaks of the weighting functions were estimated by taking the argmax of a third-order polynomial fitted to each function (the arrows in Figure 4; see Appendix C for more detail). These peaks, averaged over four subjects, are plotted against α in Figure 6a. The peak-weighted frequency tends to higher frequencies for shallower (less negative) α values;

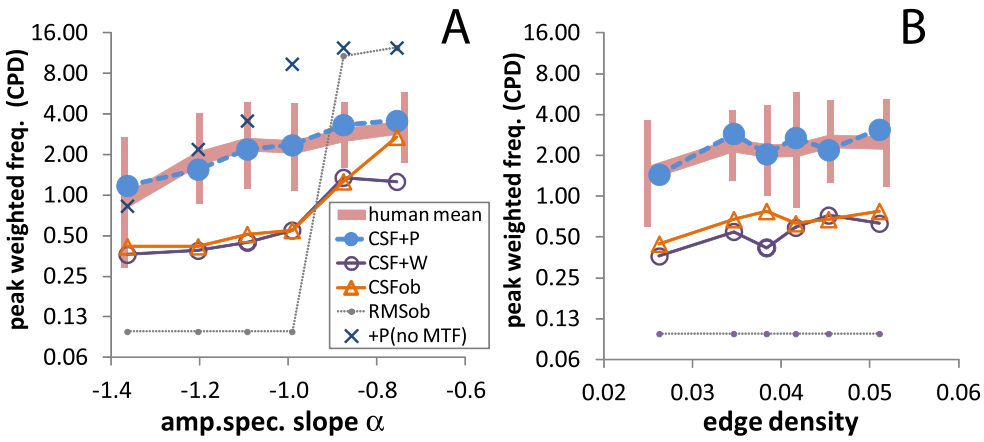

Figure 6. Peak-weighted spatial frequencies for weighting functions calculated from subsets of the data, divided into bins according to (a) the amplitude spectrum slope of the original image and (b) the edge density of the original image. Human data is averaged over the four subjects shown in Figure 4a through d and illustrated by the thick red line with the error bars corresponding to 95% confidence intervals around the means (Loftus & Masson, 1994). The model that produces simulated behavior most similar to humans is the CSF + P with the other two CSF models (CSFob and CSF + W) consistently peaking at lower frequencies. RMSob's behavior doesn't resemble the humans at all. The X symbols are for the CSF + P model with no monitor blur applied to the input images.

this makes sense because shallower α means more high-frequency contrast. Peak frequency changed on average by about 1.7 octaves over the α range shown in Figure 6a.

Figure 6a also plots peak-weighted frequencies for the models described in the previous section. CSF + P (solid round symbols) performs just like the human observers. CSFob and CSF + W weight lower frequencies than the human observers and are more affected by changes in α. RMSob fails completely: When α is less than −1.0, the RMSob chooses according to the lowest frequency contrasts, and when α is greater than −1.0, it goes by the highest frequencies. For the CSF + W and CSF + P models, performance is robust to small changes in the overall magnitude of the gain control strength; it's the way the gain control varies with spatial frequency that seems to do the trick. Having stronger gain control toward low frequencies (in the CSF + P) keeps the peak-weighted frequency higher than it would be otherwise. Meanwhile, the high-frequency contrast attenuation by the monitor MTF keeps the peaks from running up to the top of the measurement range: If the MTF is removed (X symbols in Figure 6a), the peak-weighted frequency increases steadily with spectral slope until it tops out at the highest measured frequency.

Edge density

For large, complex images like those used here, the amplitude spectrum slope says relatively little about the specific, identifiable spatial structure of an image—this being defined to a larger extent by the information in the phase spectrum (Oppenheim & Lim, 1981). A measure of the bulk structure of an image is its edge density, the amount of image area that is occupied by edges (Bex et al., 2009; Hansen & Hess, 2007); Bex et al. had proposed that local edge density might be related to the strength of contrast-gain control, which would affect judgments of perceived contrast. We measured edge density by taking the spatial average over the Canny filtered test images using the default Matlab "edge" function parameters; the distribution of values is shown in Figure 5b. Edge density and amplitude spectrum slope were correlated for our image set, but Figure 5c shows the correlation to be weak enough ($\rho = 0.447$) that the two measures can be understood to sort images into different groups. Dividing our trials into six groups by edge density and performing the same analysis as in the previous section, results were obtained as shown in Figure 6b. The dependence of peak frequency on edge density was not as clear as the dependence on spectral slope. Peak frequency increased by an average of 0.9 octaves over the edge density range shown in Figure 6b. In general, it does not seem that the edge density had a strong effect on which spatial frequencies subjects used to estimate image contrast. The performance of the different contrast models is also illustrated in Figure 6b; again the CSF + P is a close match to human performance with the CSFob and CSF + W preferring lower spatial frequencies. The RMSob averages to generally low-pass performance, choosing images based on the lowest spatial frequencies.

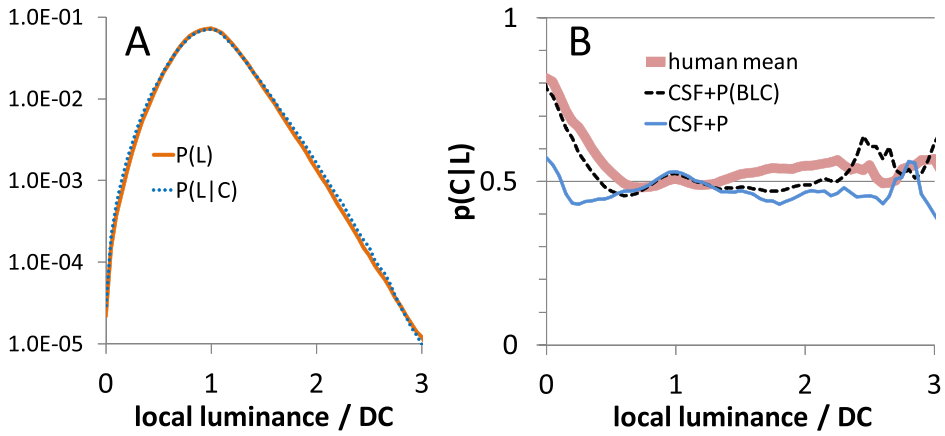

Figure 7. (a) Local luminance distributions for test images at the average peak-weighted scale, presented as the probability of a discrete picture element having the normed luminance indicated on the x-axis for all test images seen by the main human subjects of Figure 4a through d (solid orange line, P[L]) and for only the test images chosen as having higher contrast (dotted blue line, P[L|C]). Note the positive (brightward) skew characteristic of natural scenes. (b) Probability of a pixel being chosen given its luminance P(C|L) as calculated with Bayes' formula. For the human subjects (thick red line), local luminances (and, thus, images that contain them) are more likely to be chosen the darker they are from the image mean. For the CSF + P gain control model, this dark-biased behavior is seen only when contrast is computed with respect to local luminance (dashed black line) and not when computed against the DC (blue line). In (a) and (b), probabilities are shown for luminances that appeared at a rate of at least 25 pixels per test image, the lower termini of the lines in (a).

Local luminance

The magnitude of light versus dark regions was not a controlled variable in our experiment, but luminance distributions in natural scenes are not symmetric (Brady & Field, 2000) with this asymmetry dependent on both the phase and amplitude spectra. Thus, by altering the amplitude spectra, we were necessarily altering the luminance distributions, and creating room for selection. By treating the luminance distributions of our images as stimuli for selection, we attempted to recover our subjects' selection biases for luminance polarity in their judgments of image contrast.

Mean luminance of the display was allowed to vary across trials in order to maximize bandwidth for contrast display, but the two test images compared within each trial always had the same mean (DC). However, *local luminance*, understood as the mean luminance within the spatial extent of a filter of scale s (Kingdom & Moulden, 1992; Peli, 1990), did vary between the two tests because it is directly related with local contrast. We obtain our definition of local luminance from Peli (1990), who calculated local luminance at scale s and spatial position x by summing together all lower spatial frequency filter responses at the same position:

$$L(x,s) = \sum_{\phi=0}^{s-1} h_\phi * I(x). \quad (4)$$

Here, φ is the order of the filters described in Methods as used to create the experimental stimuli; $L(x,1)$ is defined as the mean luminance of the image (i.e., the local luminance relative to the coarsest contrast scale). The same calculation is involved in computing the band-limited contrast, which was used as input to the models we tested.

We analyzed local luminance by looking at luminance histograms (over all trials for each subject) for the test images at each of the eight scales availed by the filter set. In aggregating the histograms, all pixel luminances were normalized by dividing out the DC of each image. This aligned the histograms at two points: The DC was fixed at 1.0, and black was fixed to zero. This normalization scheme is similar to the reasonable assumption that most of the luminances in an image are from the same illuminant, and so if one luminance changes, all others will change by the same ratio. With this assumption the normalization can be understood as being applied to each scene's illuminant so that black is always black (zero), all scenes have similar reflectance distributions, and white is the brightest local luminance in a particular scene (Gilchrist et al., 1999), independent of both the mean and black.

All human subject data is pooled for Figure 7 as proportions of total pixels included ($\sim 3.68 \times 10^9$). Figure 7a shows the distribution of local luminances for $s = 6$—around 3 cpd—near the peak of the decision weighting functions for most of our human subjects. The skewed distribution typical of natural scenes (Brady & Field, 2000) is apparent. Consider the pixels of the luminance histograms as discrete samples of local luminance. If we treat the local luminance histograms over all test images (chosen and rejected)

as representing the probability $P(L)$ of a random pixel (at a given scale) having local luminance L; and over chosen images as representing the probability $P(L|C)$ of a random pixel having luminance L given that it was chosen (i.e., belonged to a chosen image); and if the probability that a pixel (belonging to an image) that was chosen is 0.5 (subjects had to choose between the left or right test images); then we can estimate the likelihood of a pixel being chosen given its local luminance L by using Bayes' formula: $P(C|L) = P(L|C)P(C)/P(L)$. As shown in Figure 7b (thick pink line), subjects displayed a strong bias toward choosing pixels (and thus their containing images) with darker rather than brighter local luminances. This is true despite the fact that there is a much greater dynamic range available for very bright pixels, which are several times stronger than the image mean (e.g., specular reflections or patches of sky light). Whether or not an image contained very bright regions, even several times the DC, did not particularly drive subject choices. For the CSF + P model to reproduce this aspect of human performance, it was enough that the linear filter responses were measured with respect to the local luminance (Peli, 1990) as shown by the dashed black line in Figure 7b; removing this computation removed the dark bias (blue line in Figure 7b).

Discussion

We have analyzed the results of an experiment in which human subjects compared the apparent contrast of two complex images having identical phase spectra but randomly different amplitude spectra. The humans responded as though contrast at spatial frequencies between 1 and 6 cpd contributed most to the perceived contrast of the test images. By using simulations, we showed that a CSF with converging transducers, with or without flat-weighted contrast-gain control, is not enough to explain performance in the task and demonstrated that a gain-control structure biased toward suppression of low spatial frequencies is a good predictor of performance. Studies of visual *sensitivity* have indicated the existence of such a bias through measurements of threshold changes (Bex et al., 2009; Webster & Miyahara, 1997) or by direct estimation of gain-control coefficients (Haun & Essock, 2010; Meese & Holmes, 2007).

Contrast polarity

We also presented evidence that in judging image contrast, dark regions are more important to human subjects than light regions. In judgments of the contrast of simple patterns, dark regions may count more than light regions (Chubb & Nam, 2000; Kontsevich & Tyler, 1999; Peli, 1997; Whittle, 1986). In judgments of *brightness* ("monopolar" perceived contrasts), there is clearly more perceptual gain per change in luminance for negative than for positive contrasts (Whittle, 1992). Kingdom and Whittle (1996) speculated that the minimum local luminance (for a given contrast scale) determined the response gain for contrast transduction. There is physiological evidence that the dark-sensitive regions of contrast-encoding neurons as late as the primary visual cortex are more densely innervated (Dai & Wang, 2011; Yeh, Xing, & Shapley, 2009), so contrast relative to local luminance may be driven more by dark than by bright image regions (Komban, Alonso, & Zaidi, 2011), perhaps via a gain-setting mechanism like that proposed by Kingdom and Whittle. Our model replicates human performance by computing local contrasts with respect to local luminance, so darker regions will be associated with higher contrast responses; it is also possible that dark-sensitive neurons early in the visual stream produce a response biased toward darker luminances (negative contrasts) *directly* (Komban et al., 2011; Yeh et al., 2009). However, we cannot distinguish between these alternatives with the present study. It may be common knowledge, especially in applied vision research, that the black level of a display is important to good image quality (it is obviously a selling point in display spec sheets), but we believe that we have presented the first concrete evidence that the apparent contrast of a complex, photographic image is crucially dependent on the darkness of the dark regions and not on the brightness of the light regions. With simpler stimuli, there have been numerous findings that indicate perceptual judgments related to brightness variance or contrast are driven disproportionately by dark texture elements (Chubb & Nam, 2000; Komban et al., 2011; Peli, 1997; Whittle, 1992). We have shown that if subjects are given the neutral instruction of "choose the higher contrast," they will tend to choose images with darker dark regions. We can therefore point to this aspect of our results as evidence that our subjects were, as instructed, basing their judgments of image contrast on some quantity closely related to spatial luminance variance. The main results of the experiment demonstrate biases in *scale* that contribute to those judgments.

Contrast constancy

Contrast constancy seems consistent with our subjective percepts, as it is not obvious that the phenomenal strength of images is biased toward one

scale or another. In fact, the usual rationale for contrast constancy, either suprathreshold (Georgeson & Sullivan, 1975) or near threshold (Peli, Yang, & Goldstein, 1991), is similar to the rationale for size constancy: Changes in image size or contrast could result either from optical interactions with the world or from changes in the world itself, and only the latter are of ultimate importance to perception. However, the results of our experiment are not consistent with contrast constancy in viewing of broadband images. According to the standard view, the perception of broadband patterns is essentially locked with respect to the narrowband components (Watt & Morgan, 1985); what is experienced is a *broadband* percept with a particular salience. What we have shown is that the phenomenal strength of these broadband percepts is disproportionately determined by midrange spatial frequencies. However, if we remove the display MTF from the model, its peak-weighted responses are shifted to higher frequencies (the blue X symbols in Figure 6), especially for the shallower-α images in which peak-weighted frequency reaches the limits of the measurement range (12.4 cpd). This occurs even though these test images are still subject to the camera and image interpolation MTFs. The CSF + P model thus predicts that responses to contrast in naturally viewed broadband images can be highest at very high spatial frequencies approaching (as high as an octave away from) the acuity limit.

Gain control biases

It is unclear whether or not narrowband contrast masking involves a spatial frequency bias in gain-control strength because there are few studies that have surveyed this aspect of spatial vision. Measurement of spatial frequency tuning functions for overlay masking are suggestive of stronger gain control for low spatial frequencies (Cass, Stuit, Bex, & Alais, 2009; Meese & Holmes, 2010; Wilson, McFarlane, & Phillips, 1983), but it is most significant that Meese and Holmes (2007) have shown that cross-orientation masking is stronger for high-speed targets, i.e., for patterns with low spatial frequency *and* high temporal frequency. Natural scene images modeled as a sequence of fixations tend to have increasing spectral power with increasing component speed at the same time that higher speeds correspond to lower spatial frequencies; a consequence of this (detailed in Appendix D) is that low spatial frequency is on average equivalent with high speed. So we should expect biased suppression of low spatial frequencies given the structure of our stimuli and Meese and Holmes' result. We also note that some early studies of contrast adaptation could be read as suggesting the existence of gain control mechanisms that may themselves be suppressed (Greenlee & Magnussen, 1988; Klein & Stromeyer, 1980; Nachmias, Sansbury, Vassilev, & Weber, 1973; Tolhurst, 1972). The study by Nachmias et al. is especially interesting in this respect as they seemed to show that adapting to harmonic patterns reduced or eliminated adaptation to the higher-frequency components. As support for the flat gain control hypothesis, we could point to models of blur adaptation, which have been successful at reproducing human performance with an adaptive gain control that is constant over spatial frequency (Elliott, Georgeson, & Webster, 2011; Haun & Peli, 2013a). However, although these models are structurally linked to spatial vision mechanisms, they may be describing sufficiently high-level processes that the nonlinearities of contrast transduction are less relevant than in judgments of perceived contrast.

If there is indeed a frequency bias in response suppression for broadband images, we can speculate as to the utility of such a system. The concept of response normalization is based in arguments from metabolic efficiency: If the nervous system can represent a scene just as well with less neural response, then energy is conserved and can be put to other purposes, and so such a route is likely to be taken (in development or evolution). In other words, excessive neural response is redundant and is beneficial to reduce. If we put this argument to the fact that image features like edges will stimulate neural responses simultaneously at multiple scales (Georgeson et al., 2007; Marr & Hildreth, 1980; Peli, 2002), then we might ask whether some of these responses—to, e.g., low-frequency contrasts that carry information that can be recovered from higher frequency information (Dakin & Bex, 2003; Elder, 1999; Peli, 1992)—are more expendable than others.

Keywords: contrast gain control, perceived contrast, reverse correlation, contrast contancy, natural scenes

Acknowledgments

The authors thank two anonymous reviewers for comments leading to substantive changes in the paper. Research was supported by NIH grant EY005957 to E. P. and by a grant from Analog Devices, Inc. to E. P. and A. H.

Commercial relationships: none.
Corresponding author: Andrew Morgan Haun.
Email: andrew_haun@meei.harvard.edu.
Address: Schepens Eye Research Institute, Massachussetts Eye and Ear, Harvard Medical School, Boston, MA, USA.

References

Ahumada, A., & Lovell, J. (1971). Stimulus features in signal detection. *The Journal of the Acoustical Society of America, 49,* 1751–1756.

Beard, B. L., & Ahumada, A. J. (1998). A technique to extract relevant image features for visual tasks. In Bernice E. Rogowitz & Thrasyvoulos N. Pappas (Eds.), *Proceedings of SPIE* (pp. 79–85). Human Vision and Electronic Imaging III.

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 19,* 1096–1106.

Bex, P. J., Solomon, S. G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision, 9*(10):1, 1–19, http://www.journalofvision.org/content/9/10/1, doi:10.1167/9.10.1. [PubMed] [Article]

Blakeslee, B., & McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Research, 44,* 2483–2503.

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research, 39,* 257–269.

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

Brady, N., & Field, D. J. (1995). What's constant in contrast constancy: The effects of scaling on the perceived contrast of bandpass patterns. *Vision Research, 35,* 739–756.

Brady, N., & Field, D. J. (2000). Local contrast in natural images: Normalization and coding efficiency. *Perception, 29,* 1041–1055.

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

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In J. Hornak, (Ed.), *Encyclopedia of imaging science and technology* (pp. 172–188). New York: Wiley.

Campbell, F. W., & Kulikowski, J. J. (1972). The visual evoked potential as a function of contrast of a grating pattern. *The Journal of Physiology, 222,* 345–356.

Cannon, M. W. (1985). Perceived contrast in the fovea and periphery. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 2,* 1760–1768.

Cannon, M. W. (1995). A multiple spatial filter model for suprathreshold contrast perception. In E. Peli, (Ed.), *Vision models for target detection* (pp. 88–116). Singapore: World Scientific.

Cannon, M. W., & Fullenkamp, S. C. (1991). A transducer model for contrast perception. *Vision Research, 31,* 983–998.

Cass, J., Stuit, S., Bex, P., & Alais, D. (2009). Orientation bandwidths are invariant across spatiotemporal frequency after isotropic components are removed. *Journal of Vision*, *9*(12):17, 1–14, http://www.journalofvision.org/content/9/12/17, doi:10.1167/9.12.17. [PubMed] [Article]

Chandler, D. M., & Hemami, S. S. (2007). VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing, 16,* 2284–2298.

Chen, C. C., & Tyler, C. W. (2008). Excitatory and inhibitory interaction fields of flankers revealed by contrast-masking functions. *Journal of Vision, 8*(4):10, 1–14, http://www.journalofvision.org/content/8/4/10, doi:10.1167/8.4.10. [PubMed] [Article]

Chubb, C., & Nam, J. H. (2000). Variance of high contrast textures is sensed using negative half-wave rectification. *Vision Research, 40,* 1677–1694.

Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences, USA, 86,* 9631–9635.

Dai, J., & Wang, Y. (2011). Representation of surface luminance and contrast in primary visual cortex. *Cerebral Cortex, 22,* 776–787.

Dakin, S. C., & Bex, P. J. (2003). Natural image statistics mediate brightness "filling in." *Proceedings of the Royal Society of London. Series B: Biological Sciences, 270,* 2341–2348.

Elder, J. H. (1999). Are edges incomplete? *International Journal of Computer Vision, 34,* 97–122.

Elliott, S. L., Georgeson, M. A., & Webster, M. A. (2011). Response normalization and blur adaptation: Data and multi-scale model. *Journal of Vision, 11*(2):7, 1–18, http://171.67.113.220/content/11/2/7, 10.1167/11.2.7.

Essock, E. A., DeFord, J. K., Hansen, B. C., & Sinai, M. J. (2003). Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli: A horizontal effect. *Vision Research, 43,* 1329–1335.

Essock, E. A., Haun, A. M., & Kim, Y. J. (2009). An anisotropy of orientation-tuned suppression that matches the anisotropy of typical natural scenes. *Journal of Vision, 9*(1):35, 1–15, http://www.

journalofvision.org/content/9/1/35, doi:10.1167/9.1.35. [PubMed] [Article]

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical-cells. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 4,* 2379–2394.

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.

Foley, J. M., & Chen, C. C. (1997). Analysis of the effect of pattern adaptation on pattern pedestal effects: A two-process model. *Vision Research, 37,* 2781–2788.

Garcia-Perez, M. A., & Peli, E. (2001). Luminance artifacts of cathode-ray tube displays for vision research. *Spatial Vision, 14,* 201–215.

Georgeson, M. A., May, K. A., Freeman, T. C. A., & 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, http://www.journalofvision.org/content/7/13/7, doi:10.1167/7.13.7. [PubMed] [Article]

Georgeson, M. A., & Sullivan, G. D. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *J. Physiol.-London,* 252, 627–656.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X.J., Spehar, B., Annan, V., & Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review, 106,* 795–834.

Graham, N., & Sutter, A. (2000). Normalization: Contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research, 40,* 2737–2761.

Graham, N. V. (2011). Beyond multiple pattern analyzers modeled as linear filters (as classical V1 simple cells): Useful additions of the last 25 years. *Vision Research, 51,* 1397–1430.

Greenlee, M. W., & Heitger, F. (1988). The functional-role of contrast adaptation. *Vision Research, 28,* 791–797.

Greenlee, M. W., & Magnussen, S. (1988). Interactions among spatial frequency and orientation channels adapted concurrently. *Vision Research, 28,* 1303–1310.

Hansen, B. C., & Essock, E. A. (2004). A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *Journal of Vision, 4*(12):5, 1044–1060, http://www.journalofvision.org/content/4/12/5, doi:10.1167/4.12.5. [PubMed] [Article]

Hansen, B. C., Essock, E. A., Zheng, Y., & DeFord, J. K. (2003). Perceptual anisotropies in visual processing and their relation to natural image statistics. *Network: Computation in Neural Systems, 14,* 501–526.

Hansen, B. C., & Hess, R. F. (2007). Structural sparseness and spatial phase alignment in natural scenes. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 24,* 1873–1885.

Hansen, B. C., & Hess, R. F. (2012). On the effectiveness of noise masks: Naturalistic vs. unnaturalistic image statistics. *Vision Research, 60,* 101–113.

Haun, A. M., & Essock, E. A. (2010). Contrast sensitivity for oriented patterns in 1/f noise: Contrast response and the horizontal effect. *Journal of Vision, 10*(10):1, 1–21, http://www.journalofvision.org/content/10/10/1, doi:10.1167/10.10.1. [PubMed] [Article]

Haun, A. M., & Peli, E. (2013a). Adaptation to blurred and sharpened video. *Journal of Vision, 13*(8):12, 1–14, http://www.journalofvision.org/content/13/8/12, doi:10.1167/13.8.12. [PubMed] [Article]

Haun, A. M., & Peli, E. (2013b). Is image quality a function of contrast perception? In *IS&T/SPIE electronic imaging* (pp. 86510C–86510C).

Haynes, J. D., Roth, G., Stadler, M., & Heinze, H. J. (2003). Neuromagnetic correlates of perceived contrast in primary visual cortex. *Journal of Neurophysiology, 89,* 2655–2666.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7,* 498–504.

Huang, P.-C., Maehara, G., May, K. A., & Hess, R. F. (2012). Pattern masking: The importance of remote spatial frequencies and their phase alignment. *Journal of Vision, 12*(2):14, 1–13, http://www.journalofvision.org/content/12/2/14, doi:10.1167/12.2.14. [PubMed] [Article]

Kim, Y. J., Haun, A. M., & Essock, E. A. (2010). The horizontal effect in suppression: Anisotropic overlay and surround suppression at high and low speeds. *Vision Research, 50,* 838–849.

Kingdom, F. A., & Moulden, B. (1992). A multi-channel approach to brightness coding. *Vision Research, 32,* 1565–1582.

Kingdom, F. A. A., & Whittle, P. (1996). Contrast discrimination at high contrasts reveals the influence of local light adaptation on contrast processing. *Vision Research, 36,* 817–829.

Klein, S., & Stromeyer, C. F. (1980). On inhibition between spatial frequency channels: Adaptation to complex gratings. *Vision Research, 20,* 459–466.

Komban, S. J., Alonso, J. M., & Zaidi, Q. (2011). Darks are processed faster than lights. *Journal of Neuroscience, 31,* 8654–8658.

Kontsevich, L. L., & Tyler, C. W. (1999). Non-linearities of near-threshold contrast transduction. *Vision Research, 39,* 1869–1880.

Kontsevich, L. L., & Tyler, C. W. (2013). A simpler structure for local spatial channels revealed by sustained perifoveal stimuli. *Journal of Vision, 13*(1):22, 1–12, http://www.journalofvision.org/content/13/1/22, doi:10.1167/13.1.22. [PubMed] [Article]

Kulikowski, J. J. (1976). Effective contrast constancy and linearity of contrast sensation. *Vision Research, 16,* 1419–1431.

Kwon, M. Y., Legge, G. E., Fang, F., Cheong, A. M. Y., & He, S. (2009). Adaptive changes in visual cortex following prolonged contrast reduction. *Journal of Vision, 9*(2):20, 1–16, http://www.journalofvision.org/content/9/2/20, doi:10.1167/9.2.20. [PubMed] [Article]

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

Levi, D. M., Klein, S. A., & Chen, I. N. (2005). What is the signal in noise? *Vision Research*, *45,* 1835–1846.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence-intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

Lubin, J. (1995). A visual discrimination model for imaging system design and evaluation. In E. Peli (Ed.), *Vision models for target detection* (pp. 245–357). Singapore: World Scientific.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *207*, 187–217.

Meese, T. S., & Baker, D. H. (2011). Contrast summation across eyes and space is revealed along the entire dipper function by a "Swiss cheese" stimulus. *Journal of Vision, 11*(1):23, 1–23, http://www.journalofvision.org/content/11/1/23, doi:10.1167/11.1.23. [PubMed] [Article]

Meese, T. S., & Hess, R. F. (2004). Low spatial frequencies are suppressively masked across spatial scale, orientation, field position, and eye of origin. *Journal of Vision, 4*(10):2, 843–859, http://www.journalofvision.org/content/4/10/2, doi:10.1167/4.10.2. [PubMed] [Article]

Meese, T. S., & Holmes, D. J. (2007). Spatial and temporal dependencies of cross-orientation suppression in human vision. *Proceedings of the Royal Society B: Biological Sciences, 274,* 127–136.

Meese, T. S., & Holmes, D. J. (2010). Orientation masking and cross-orientation suppression (XOS): Implications for estimates of filter bandwidth. *Journal of Vision, 10*(12):9, 1–20, http://www.journalofvision.org/content/10/12/9, doi:10.1167/10.12.9. [PubMed] [Article]

Nachmias, J., Sansbury, R., Vassilev, A., & Weber, A. (1973). Adaptation to square-wave gratings: In search of the elusive third harmonic. *Vision Research, 13,* 1335–1342.

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

Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 7,* 2032–2040.

Peli, E. (1992). Perception and interpretation of high-pass filtered images. *Optical Engineering, 31,* 74–81.

Peli, E. (1997). In search of a contrast metric: Matching the perceived contrast of Gabor patches at different phases and bandwidths. *Vision Research, 37,* 3217–3224.

Peli, E. (2002). Feature detection algorithm based on a visual system model. *Proceedings of the IEEE, 90,* 78–93.

Peli, E., Yang, J., & Goldstein, R. B. (1991). Image invariance with changes in size: The role of peripheral contrast thresholds. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 8,* 1762–1774.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31,* 1337–1350.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems, 10,* 341–350.

Ross, J., & Speed, H. D. (1991). Contrast adaptation and contrast masking in human vision. *Proceedings of the Royal Society B-Biological Sciences, 246,* 61–70.

Schofield, A. J., & Georgeson, M. A. (2003). Sensitivity to contrast modulation: The spatial frequency dependence of second-order vision. *Vision Research, 43,* 243–259.

Swanson, W. H., Wilson, H. R., & Giese, S. C. (1984).

Contrast matching data predicted from contrast increment thresholds. *Vision Research,* 24, 63–75.

Taylor, C. P., Bennett, P. J., & Sekuler, A. B. (2009). Spatial frequency summation in visual noise. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 26,* B84–B93.

Teo, P. C., & Heeger, D. J. (1994). Perceptual image distortion. In *Image Processing* (pp. 982–986). IEEE International Conference, doi:10.1109/ICIP.1994.413502.

To, M. P. S., Lovell, P. G., Troscianko, T., & Tolhurst, D. J. (2010). Perception of suprathreshold naturalistic changes in colored natural images. *Journal of Vision, 10*(4):12, 1–22, http://www.journalofvision.org/content/10/4/12, doi:10.1167/10.4.12. [PubMed] [Article]

Tolhurst, D. J. (1972). Adaptation to square-wave gratings: Inhibition between spatial frequency channels in the human visual system. *The Journal of Physiology, 226,* 231–248.

Tyler, C. W. (1997). Colour bit-stealing to enhance the luminance resolution of digital displays on a single pixel basis. *Spatial Vision, 10,* 369–377.

Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 14,* 2379–2391.

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

Webster, M. A., & Miyahara, E. (1997). Contrast adaptation and the spatial structure of natural images. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 14,* 2355–2366.

Whittle, P. (1986). Increments and decrements: Luminance discrimination. *Vision Research, 26,* 1677–1691.

Whittle, P. (1992). Brightness, discriminability and the "crispening effect." *Vision Research, 32,* 1493–1507.

Wilson, H. R., & Humanski, R. (1993). Spatial-frequency adaptation and contrast gain-control. *Vision Research, 33,* 1133–1149.

Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial-frequency tuning of orientation selective units estimated by oblique masking. *Vision Research, 23,* 873–882.

Yang, J., Qi, X. F., & Makous, W. (1995). Zero-frequency masking and a model of contrast sensitivity. *Vision Research, 35,* 1965–1978.

Yeh, C. I., Xing, D. J., & Shapley, R. M. (2009). "Black" responses dominate macaque primary visual cortex V1. *Journal of Neuroscience, 29,* 11753–11760.

Appendix A: Modeling perceived contrast

Models of contrast perception

In the following sections, we describe the model components that are both sufficient to emulate human performance in our experiment and consistent with the broader contrast-perception literature. The first component is a CSF. The second is a nonlinear contrast transducer of the familiar form, producing dipper-shaped contrast discrimination functions (Legge & Foley, 1980). The third is a spatial-frequency bias in contrast gain-control strength (Haun & Essock, 2010). The fourth is that the contrast input to the model is relative to local luminance. Except for the third component, all of these are uncontroversial features of spatial vision.

CSF and nonlinear transducer

The CSF is usually not considered to contribute much to suprathreshold contrast perception (Brady & Field, 1995; Georgeson & Sullivan, 1975) although some studies suggest that component contrasts are not perceptually constant with scale (Bex et al., 2009; Haun & Essock, 2010) or even that they are consistent with the blank-adapted CSF (Bex & Makous, 2002; Kulikowski, 1976). Models of broadband contrast sensitivity, especially those used in the study of image quality and discrimination (Chandler & Hemami, 2007; Lubin, 1995; To, Lovell, Troscianko, & Tolhurst, 2010), use human CSFs to set thresholds for nonlinear contrast transducers that converge at high contrasts. Despite this high contrast constancy, these sorts of models would still be expected to give larger overall responses to contrasts near the CSF peak (lower threshold) because most image contrasts are low relative to the thresholds (Brady & Field, 2000) and thus will indeed produce, in the aggregate, CSF-modulated responses.

Cannon's spatial model of perceived contrast (Cannon, 1995; Cannon & Fullenkamp, 1991) is a good starting point for any model of perceived contrast of spatially extended patterns. It incorporates the familiar nonlinear transducers, compulsorily combined over frequency (through Minkowski summation), which can also be used to predict contrast discrimination and detection (Swanson et al., 1984). Simplifying somewhat, for a given spatial location x, Cannon's model is

expressed as

$$R(x) = \left(\sum_f \left[r \frac{|C_f(x)|^{p+q}}{z_f^p + |C_f(x)|^p} \right]^M \right)^{1/M} \quad \text{(A1)}$$

where C is a linear, band-pass measure of contrast, and R is perceived contrast at image location x. Here, z sets the transducer threshold and is dependent on frequency f; r scales the transducer, so it can be equated with an arbitrary performance measure (d' in our case; we set $r = 40$), and is constant with frequency (which, in itself, implies a deblurring operation of the sort described by Georgeson & Sullivan (1975) and Brady & Field (1995); p and q set the rate of response change with contrast change (we set these to typical values of 2.0 and 0.4, respectively; cf. Legge & Foley, 1980); and M values greater than 1.0 bias the response R toward stronger frequency-specific (winner-take-all) responses. We set M equal to 4.0, as per Cannon & Fullenkamp (1991) and Swanson et al. (1984). The input to the function is the rectified response $|C|$ of a cosine-phase filter to the image at position x (we used the same nonoriented filters as were used to generate the test stimuli, cf. Methods). In Cannon's original model, the z term was partly dependent, via spatial normalization, on the area of the stimulus, and the observer's judgments were based on a MAX operator over $R(x)$. We instead made the final estimate another Minkowski sum over space with the same exponent M, reflecting the disproportionate effects that high local contrasts would likely have on observer judgments:

$$\mathbf{R} = \frac{1}{N} \left(\sum_{x=1}^{N} R(x)^M \right)^{1/M} \quad \text{(A2)}$$

Here, N is the number of pixels (filter locations and discrete responses R) in the test image. For experiment simulation, a random value with unit Gaussian variance was added to each computed **R** value. The human contrast thresholds were used to set the z values of Equation A1 by solving for the following (a rearrangement of the transducer of Equation A1 at threshold):

$$z_f^p = \frac{r}{d'} \cdot {t_f}^{p+q} - {t_f}^p \quad \text{(A3)}$$

where t was the value of a smooth function (Haun & Essock, 2010; Yang, Qi, & Makous, 1995) fitted to the average thresholds for our subjects, which allowed for different filter frequencies to be used if desired. d' here was set to 1.0, near the threshold level sought in the human CSF measurement (Appendix B).

We did not include any variation in sensitivity with visual field location. Because the observers were free to foveate the entire display, we assume that our results mainly reflect what was seen in the central visual field, and so the spatial summation of Equation A2 can be considered a spatiotemporal summation over many eye movements. On a related point, we interpret the spatial M-norm of Equation A2 to reflect the disproportionate impact of strong local contrasts on observer judgments, including the higher likelihood that these contrasts will be foveated (Reinagel & Zador, 1999), rather than a lower-level pooling mechanism, which is unlikely to take the form of a high-M norm (Meese & Baker, 2011).

The simulated observer chose the test image in each trial that produced the larger **R** value.

Pattern masking/contrast gain control

Adaptation and suppression are ubiquitous processes in spatial vision that affect contrast perception and have been included in successful models of image discriminability (Teo & Heeger, 1994; To et al., 2010; Watson & Solomon, 1997). Measurements of contrast sensitivity at different spatial frequencies during viewing of broadband patterns (Haun & Essock, 2010), after adaptation to the same (Bex et al., 2009; Webster & Miyahara, 1997), or with narrowband cross-oriented masks (Meese & Hess, 2004) all suggest that masking and/or adaptation processes are stronger toward lower spatial frequencies. The functional form of this bias has not been described, but estimates by Haun and Essock (2010) are generally consistent with adaptation strength inversely proportional to spatial frequency if the adaptation is formulated as an increase in transducer threshold (the denominator in Equation A1) dependent on stimulus contrast (Foley, 1994; Foley & Chen, 1997):

$$R(x) = \left(\sum_f \left[r \frac{|C(x)|^{p+q}}{z_f^p + |C(x)|^p + w_f |C(x)|^p} \right]^M \right)^{1/M} \quad \text{(A4)}$$

Here, w_f is the gain-control weighting term, analogous to w_i in Equation 2. For model CSF + P, w_f was set equal to $a/f^{I/p}$, similar to the dependence of gain control on stimulus *speed* (temporal divided by spatial frequency) observed by Meese and Holmes (2007); see Appendix D for more discussion on this point. The general configuration is similar to Foley's (1994) model as expressed in Equation 2 except that the filters here are not oriented, and the source and target of masking are always identical. So wherever there is contrast, there will be masking. This is a very simplified representation. There are suppressive interactions between distant spatial frequencies (Foley, 1994; Greenlee & Magnussen, 1988; Huang et al., 2012; Meese & Hess, 2004), different orientations (Foley, 1994; Meese & Holmes, 2007), and different

spatial locations (Chen & Tyler, 2008); contrast adaptation can be considered a type of self-suppression (Foley & Chen, 1997; Greenlee & Heitger, 1988; Wilson & Humanski, 1993). Any given image contrast will mostly likely be influenced by one or more of these types of masking, all of which have functional similarities, with spatial suppression being a possible exception (Chen & Tyler, 2008). The formulation of Equation A4 is intended to represent an average or agglomeration of these different processes.

This arrangement, with a set to 0.5 (the gain control scheme, including the constant value, was adjusted in coarse steps to produce the observed fit to the human data), is sufficient to reproduce the general form of results obtained by the psychophysical studies cited above. This model is the model $CSF + P$ referred to in the main text. Model $CSF + W$ was similar except that the masking coefficient w was fixed at 0.5, making the effect of gain control constant over all frequencies and giving CSF + W a similar overall response magnitude as CSF + P. With w set to zero, we have model *CSFob*.

Contrast polarity

By taking the filter responses directly and feeding them into Equation A4, we are implicitly measuring contrast with respect to the image DC. Light adaptation is a retinotopically local process, so this is an inaccurate means of measuring contrast. By taking the local luminance into account when calculating the filter responses, we can make more physiologically and perceptually meaningful measures of image contrast: the band-limited contrast (Peli, 1990). With the definition of local luminance $L(s)$ given in Equation 4 (in the main text), we adopted the following definition of band-limited contrast (BLC) **C**:

$$\mathbf{C}(x,s) = \frac{|C(x,s)|}{\max[L(x,s)/L(x,0), \varepsilon]}. \quad \text{(A5)}$$

Here, ε is a small value to prevent division by zero. Recall that s refers to a given contrast scale, so contrasts at scale s will be measured against the sum of all frequencies at lower scales, relative to the DC. This means that if the local luminance is equal to the DC, contrast will be as with the original filter response $|C|$. Importantly, if $L(s)$ is higher than the DC, the contrast response will be *less*, and if $L(s)$ is lower than the DC, the contrast response will be *greater*. A similar relationship between contrast perception and local luminance was described by Whittle (1986, 1992), who noted that the local *minimum* luminance seemed to exert an important influence on perceptual gain for increments or decrements in luminance, thus determining judgments of brightness and darkness.

An alternative approach to accounting for the luminance polarity effect would be to implement an imbalance in the gains of negative (dark) versus positive (light) polarity filters. This would be in line with neurophysiological evidence that suggests a numerical bias for off-center neurons in the primary visual cortex (Yeh et al., 2009). To simulate such an imbalance, we increased negative $C(x)$ responses by 50% and decreased positive $C(x)$ responses by the same amount. This manipulation had no effect on the luminance-weighting functions (not shown). The reason for this is that the *within-band* distribution of positive and negative responses is symmetric—determined by the symmetry of the filters—so when there are negative responses in one location, there will be positive responses of similar magnitude elsewhere; the asymmetry of natural luminance distributions is the result of phase correspondences over multiple scales, which are not encoded in the individual contrast bands. A simple imbalance in the responsiveness of negative versus positive filters is not enough to explain human performance in this task. Sensitivity to local luminance must be incorporated at some point. The reason that using band-limited contrast emulates the human performance is that it incorporates local luminance into contrast calculations in such a way that lower luminances are associated with higher contrasts so that suppressed low-frequency contrasts are still represented, in a way, in responses to higher frequencies. Although most primary visual cortex neurons are relatively insensitive to local luminance, it has been shown that those cortical neurons that do encode local luminance mostly encode darkness and that most of these are still predominantly driven by contrast (Dai & Wang, 2011). We speculate that if our model—which falls a bit short of human performance as shown in Figure 7—used filters sensitive to local luminance, rather than the precisely band-pass filters used, we could then adjust the strength of this sensitivity in negative versus positive filters to obtain results even more similar to the human data.

Appendix B: Observer CSFs and monitor MTF

Threshold measurement

To calibrate the perceived contrast models, we measured our observers' thresholds for band-pass versions of the test images. The same raised-cosine filters used in the main experimental manipulation were used to produce the test images, which were normalized so that their RMS contrast could be

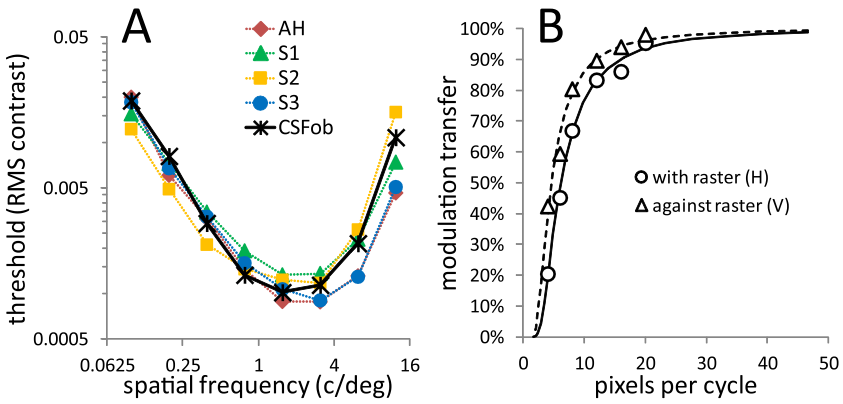

Figure B1. Empirical functions used to calibrate the models described in the text. (a) CSFs for four human observers (colored markers and lines) and for the simulated observers (black Xs, black line). (b) The modulation transfer function for the experiment display. Symbols are estimated attenuation measurements; lines are Gaussian fits (in units of spatial frequency).

directly specified. Display conditions (distance, monitor calibration, etc.) were similar to the main experiment except that display mean luminance was kept constant at 39 cd/m^2. Thresholds were measured with a three-down one-up staircase varying band contrast in 1 dB steps, where subjects indicated which side of the screen contained a target image. Trials were untimed (as with the main experiment) with numeric cues at the center of the display to inform subjects as to what scale they were looking for on a given trial (numbers ranging from one to eight for the eight bands; the cues were explained with high-contrast demo stimuli before the task was run). Sixteen staircases were randomly interleaved in blocks of 512 trials (32 trials each), eight for each band on each side of the screen. Threshold was defined as the contrast yielding 81% correct and was estimated by fitting a logistic function to the trial data. Each subject ran either three or four blocks of the detection task for a total of 192 or 256 trials per threshold estimate. The four subjects' CSFs (Figure B1a) were averaged and used (via a fitted function) to establish the simulated observer's thresholds as described in Appendix A. Model thresholds are plotted in Figure B1a (black asterisks) and were obtained by finding the band RMS contrasts that, on average, brought Equation A4 to a value of 1.0 (with the masking term set to zero—effectively the CSFob of Equation A2 was used).

MTF measurement

We measured the voltage modulation of a photodiode response (OTR-3, Messtechnik) to square wave gratings of variable spatial frequency drifting slowly (1 px/s at the 100-Hz refresh rate) past a 1 mm slit aperture, emulating the method prescribed by Pelli and Zhang (1991) and Brainard, Pelli, and Robson (2002). Measurements were made with the grating/slit oriented with or against the raster. Because the slit sat over 2.67 pixels, the measured modulation was divided by values computed for an ideal system with no blur and the same pixel and slit dimensions as the physical system. The ideal system was a 2.67-pixel aperture convolved with the test square waves, which produced a blurred wave; the RMS power of the (convolved) output wave was divided by the RMS power of the input wave. These ratios represent the blur of the measurement technique. The RMS power of the photodiode response at each spatial frequency was divided by the technique ratios, which were fit with a Gaussian function of spatial frequency; this was then scaled to unity at the DC (Figure B1b). The outer product of the with- and against-raster functions (the Fourier transform of the pixel spread function) was used to attenuate the amplitude spectra of the test images used in the experiment simulations. By this procedure, we only took into account the linear blur of the display and ignored more complex nonlinear CRT artifacts (Garcia-Perez & Peli, 2001).

Appendix C: Peaks of the weighting functions

For the analyses shown in Figure 6, we interpolated the weighting-function peaks by fitting each function with a third-order polynomial. Figure C1 shows polynomial fits to the weighting functions for four subjects and one model observer (rows) whose trials were divided into six groups according to the α value for the test images in each trial. The maxima of the polynomials within the bounds set by the highest and

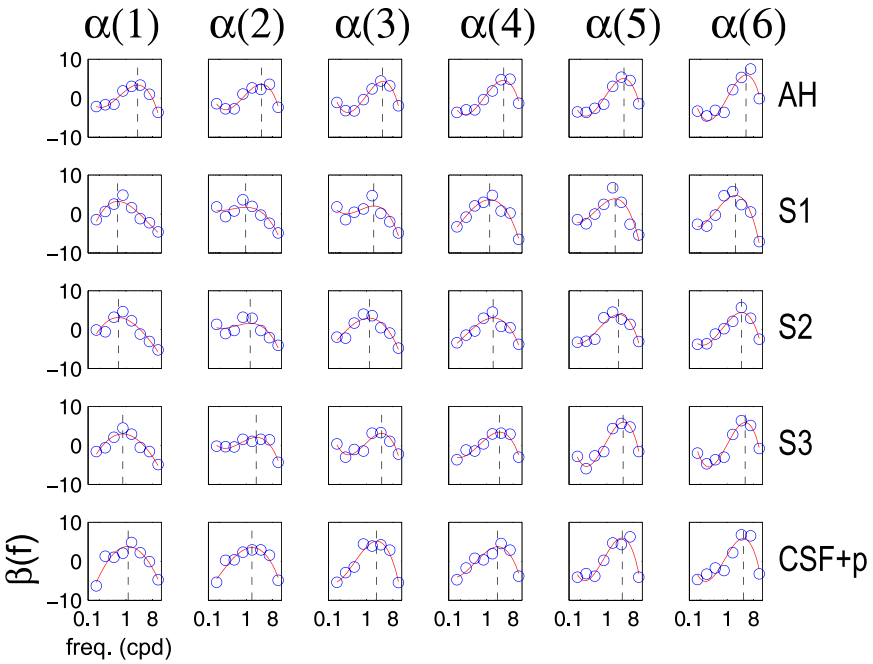

Figure C1. Decision-weighting functions (blue symbols), fitted polynomials (red lines), and peak-weighted frequencies (dashed black lines) for the four main subjects of Figure 3 and the most successful model we tested, CSF + P. These are the data that produced the peak-weighted frequencies plotted in Figure 6.

lowest test frequencies were used to interpolate the peak-weighted frequencies plotted in Figure 6. Any number of other calculations might have been used to estimate the peaks of the weighting functions, but we think that readers will agree that other methods would have yielded similar estimates. At this level of detail, the similarity between human and model performance is apparent. In one respect, there is an obvious divergence: The human weighting functions tend to deflect toward zero at the lowest spatial frequency band whereas the model observer does not. This may reflect the absence (Kontsevich & Tyler, 2013) or some fundamentally different qualities (Wilson et al., 1983) of very low frequency channels in the human observers.

Appendix D: Speed spectra of natural scene sequences

We asserted in the Discussion that a sequence of fixations of a static natural image will yield a spatiotemporal power spectrum in which speed is inversely proportional to spatial frequency and independent of temporal frequency. In combination with Meese & Holmes' (2007) finding that (for simple grating stimuli) contrast gain–control strength increased proportionally to the square root of stimulus speed, this seems to be enough to explain the gain-control pattern implicated in the results of our experiment. To examine the spatiotemporal structure of sequentially fixated static imagery similar to what we used in our experiments, we used as a "fixation source" randomly selected scenes from the original experiment with a resolution of 1024 × 1024 pixels. We modeled periodic fixation as sequences of 16 128 × 128-pixel patches drawn from random locations within the source image. Each fixation patch was repeated in 16 consecutive "frames," so the sequence consisted of a stack of 256 patches:

$$\left\{ \begin{bmatrix} p_{1,1} & p_{1,\cdots} & p_{1,16} \end{bmatrix} \quad \begin{bmatrix} p_{\cdots,1} & p_{\cdots,\cdots} & p_{\cdots,16} \end{bmatrix} \right.$$
$$\left. \begin{bmatrix} p_{16,1} & p_{16,\cdots} & p_{16,16} \end{bmatrix} \right\}$$

Each 256-frame sequence was scaled to a 4-s period, so there were four "fixations" per second (close to the typical duration of a fixation while free viewing a natural scene; Henderson, 2003), and the patches were scaled to 2° across (H/V).

We then took the spatiotemporal power spectrum of the sequence (using the Matlab fftn( ) function) and filtered this with an array of spatiotemporal filters with Gaussian orientation profile (width at a half height of 45° at the four cardinal orientations), log Gaussian spatial frequency profile (1 octave width, octave spaced from 1 to 32 cpd), and log Gaussian temporal frequency profile (1 octave width, spaced from 1 to 16 Hz). The sums of the filtered power spectra—proportional to the RMS contrast of the filtered sequences—

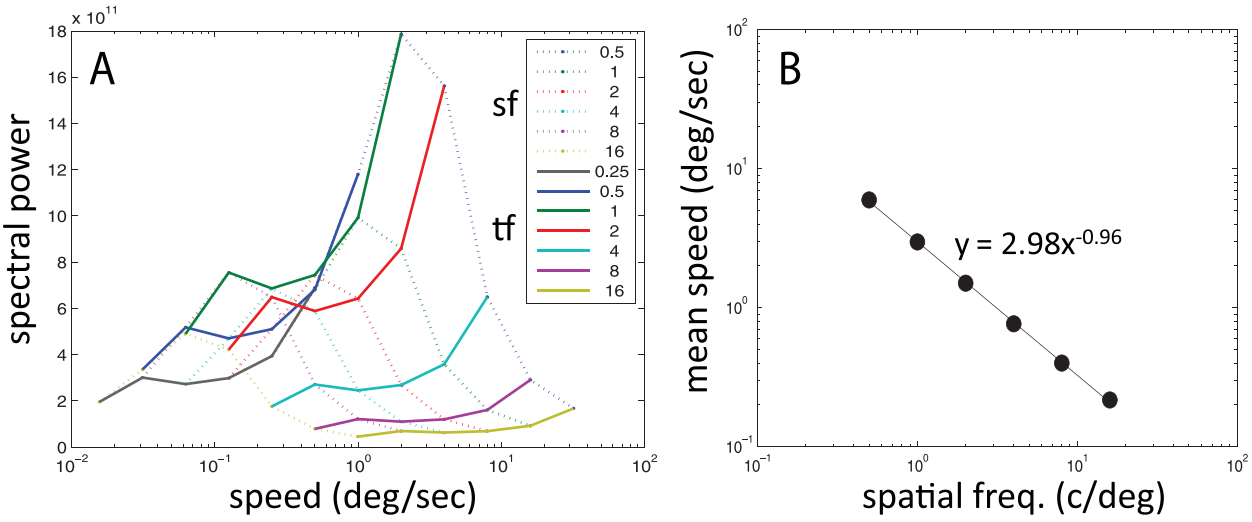

Figure D1. (a) Power/speed plots for 128 natural scene patch sequences (averaged after filtering). The x-axis is speed, the ratio of the peak spatial and temporal frequencies of a spatiotemporal Gaussian filter. The y-axis is the sum of the filtered power spectrum of the model fixation sequence. Solid lines are for speeds with the same temporal frequency (for each line color, according to the legend) but variable spatial frequency, and the dotted lines are for speeds with the same spatial frequency but variable temporal frequency. (b) The average speed for each spatial frequency shown in the left panel. The y-axis shows the average speed of the dotted lines in panel (a), weighted by spectral power.

averaged over orientation are plotted in Figure D1a. The dotted lines represent speeds with constant spatial frequency but changing temporal frequency (speed increases with temporal frequency when spatial frequency is constant, so for each dotted line, temporal frequency increases from left to right); the solid lines represent speeds with constant temporal frequency but changing spatial frequency (speed decreases with increasing spatial frequency, so for each solid line, spatial frequency decreases from left to right). There is clearly a progression from less power at low speeds to more power at high speeds, especially along the iso-temporal–frequency (solid) lines. A normalization system, considered within a given temporal frequency band, would see higher power at and should exert more suppression on lower spatial frequencies.

To clarify what is happening in Figure D1, we can collapse over temporal frequency, showing directly the relationship between speed and spatial frequency. We used the total power for each spatiotemporal filter to calculate a weighted mean over temporal frequency (with f and η representing spatial and temporal frequency, respectively):

$$V(f) = \frac{\sum_{\eta} \frac{\eta}{f} P(\eta, f)}{\sum_{\eta} P(\eta, f)}$$

It is not hard to see that if the average temporal frequency is flat over spatial frequency, the average speed will be inversely proportional to spatial frequency; because sequential views of static imagery contribute the same temporal modulation to every frequency, we should expect this to be so. Figure D1b shows that $V(s)$ is exactly the inverse of spatial frequency times a constant. Given this equivalence between spatial frequency and speed, the gain control–weighting scheme we used in Equation A4 is nearly identical to that found by Meese and Holmes (2007). Their masking term w was equated to around .04 $\times$ $V^{0.5}$, and in equivalent terms, our masking term of $0.5/f^{1/p}$ comes out to around $0.29 \times V^{0.5}$ (by combining our masking term with the power function of Figure D1b). This difference in strength (almost an order of magnitude) may reflect the contribution of multiple suppressive influences—all with the same speed bias—to the gain control mechanism.