The relation of phase noise and luminance contrast to overt attention in complex visual stimuli

Wolfgang Einhäuser Division of Biology, California Institute of Technology, Pasadena, CA, USA

Ueli Rutishauser Computation and Neural Systems, California Institute of Technology, Pasadena, CA, USA

E. Paxon Frady Computation and Neural Systems, California Institute of Technology, Pasadena, CA, USA

Swantje Nadler Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Peter König Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

Christof Koch Division of Biology, Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA, USA

Models of attention are typically based on difference maps in low-level features but neglect higher order stimulus structure. To what extent does higher order statistics affect human attention in natural stimuli? We recorded eye movements while observers viewed unmodified and modified images of natural scenes. Modifications included contrast modulations (resulting in changes to first- and second-order statistics), as well as the addition of noise to the Fourier phase (resulting in changes to higher order statistics). We have the following findings: (1) Subjects' interpretation of a stimulus as a "natural" depiction of an outdoor scene depends on higher order statistics in a highly nonlinear, categorical fashion. (2) Confirming previous findings, contrast is elevated at fixated locations for a variety of stimulus categories. In addition, we find that the size of this elevation depends on higher order statistics and reduces with increasing phase noise. (3) Global modulations of contrast bias eye position toward high contrasts, consistent with a linear effect of contrast on fixation probability. This bias is independent of phase noise. (4) Small patches of locally decreased contrast repel eye position less than large patches of the same aggregate area, irrespective of phase noise. Our findings provide evidence that deviations from surrounding statistics, rather than contrast per se, underlie the well-established relation of contrast to fixation.

Keywords: attention, eye movements, saliency, phase noise, higher order statistics, natural scenes, fractals

Introduction

When viewing complex stimuli, human observers sequentially shift their attention (James, 1890). In natural vision, shifts in eye position correlate tightly with such attentional shifts (Rizzolatti, Riggio, Dascola, & Umilta, 1987). Various factors guide this "overt" attention, such as the features of the stimulus, the observer's experience, and the task (Buswell, 1935; Yarbus, 1967). Most models of human attention focus on the former "bottom–up" signals, resting upon the concept of a "saliency map" (Koch & Ullman, 1985). According to this scheme, stimuli are processed in multiple independent feature channels, local differences ("contrasts") in these channels are summed, and the activity in the resulting saliency map reflects the probability of a location to be attended. Various implementations of the saliency-map scheme predict human fixation behavior in natural scenes better than chance (Itti & Koch, 2000; Parkhurst, Law, & Niebur, 2002; Peters, Iyer, Itti, & Koch, 2005; Tatler, Baddeley, & Gilchrist, 2005). One of the model's features—luminance contrast—is elevated at fixation, as compared with random locations (Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Reinagel & Zador, 1999). This effect, however, is contingent on correcting for a general fixation bias toward the image center (Mannan, Ruddock, & Wooding, 1996, 1997) or on restricting analysis to certain spatial frequencies (Einhäuser & König, 2003; Tatler et al., 2005). In addition, this correlation does not imply a causal contribution of luminance contrast to fixation but rather reflects the correlation of both with a higher order stimulus property (Einhäuser &

doi: 10.1167/6.11.1 Received September 2, 2005; published October 9, 2006 ISSN 1534-7362

König, 2003). This raises the question to what extent the effect of a low-level feature, such as luminance contrast, on attention depends on higher order stimulus statistics.

Several studies directly measure the effect of higher order statistics on human attention. By analyzing bispectral densities, Krieger et al. (2000) find higher order "structural differences" between fixated and nonfixated regions. These authors propose two-dimensional image properties, like curves, edges, spots, and so forth, to underlie the selection of fixation points. Privitera, Fujita, Chernyak, and Stark (2005) not only identify several generic geometric kernels that are good predictors of fixated regions but also stress that their results are only valid for "generic" images—in their case, landscapes, interiors, and object collections—and in the absence of a specific task. In earlier work (Privitera & Stark, 2000), the same authors had compared 10 different algorithms for predicting fixation locations in a variety of images. The performance of any algorithm depends largely on the image category: For example, while contrast is a good predictor in "terrain" scenes, symmetry is more important in artistic paintings. Consequently, results on attention in "natural" scenes have to be probed for such category dependence.

The averaged amplitude spectra of natural scenes tend to follow a $1/f$ law (Betsch, Einhäuser, Körding, & König, 2004; Field, 1987; Ruderman & Bialek, 1994; Torralba & Oliva, 2003; van der Schaaf & van Hateren, 1996). Most information on the content of a specific natural scene, however, seems to be contained in its phase spectrum: When mixing the amplitude spectrum of one image with the phase spectrum of another, the image contributing the phase dominates the perception of the mixture (Oppenheim & Lim, 1981). Adding noise to the phase of a stimulus modulates responses in the visual cortex of macaque monkeys (Rainer, Augath, Trinath, & Logothetis, 2001, but see Dakin, Hess, Ledgeway, & Achtman, 2002) and also impairs their performance in a memory task (Rainer, Lee, & Logothetis, 2004). High levels of phase noise also impair human performance in a rapid categorization task, but some category information is retained in the amplitude spectrum (Wichmann, Braun, & Gegenfurtner, 2006). In the context of overt attention, monkeys are less likely to fixate areas of the stimulus that are locally deprived of phase information, which can be compensated for by a local increase in luminance contrast (Kayser, Nielsen, & Logothetis, 2006). Because monkey and human fixations are affected differently by subtle local changes of luminance contrast (Einhäuser, Kruse, Hoffmann, & König, 2006), it is unclear whether this result transfers to human observers.

We investigate how higher order stimulus statistics interact with a first-order feature in guiding human overt attention. To do so, we test the following questions:

1. Are higher order stimulus statistics needed for the subjective perception of an outdoor scene as natural?
2. Does the elevation of luminance contrast at fixations depend on higher order statistics or stimulus category?
3. Do large-scale variations of luminance contrast bias attention irrespective of higher order statistics?
4. Do local variations of luminance contrast have similar effects as global changes?

We address these questions by measuring eye movements of human observers while they view statistically modified images of outdoor scenes, man-made objects, human faces, and fractals.

Methods

Stimuli

We performed four separate experiments using grayscale images. All experiments used outdoor images based on the Zurich Natural Image Database (Einhäuser et al., 2006; http://www.klab.caltech.edu/~wet/ZurichNatDB.tar.gz), which contain none or very few man-made objects. In Experiment 1, three additional categories were used: "man-made objects" from the McGill calibrated color image database (Olmos A. and Kingdom F. A. A.), "fractals" from the chaotic n-space network database (http://www.cnspace.net/html/fractals_gallery.html), and frontal views of 16 different faces taken with a Sony DSC-V1 cybershot camera (Sony, Tokyo, Japan) under controlled lighting conditions (Figure 1a). In Experiments 1 and 4, stimuli were used at a resolution of 1,024 × 768 pixels and 8-bit grayscale; in Experiments 2 and 3, images were centrally cropped to 768 × 768 pixels.

Phase noise

To manipulate the higher order statistics of a stimulus, we modified its phase spectrum. The amplitude spectrum was unchanged. We transformed the original images to Fourier space, added noise to the phases, and transformed the combination of amplitude and phase back into image space. The additive noise was drawn from a normal distribution of standard deviation σ (Experiments 2 and 3) or a symmetric uniform distribution of width η (Experiments 1 and 4) and zero mean. To minimize the effects on overall contrast in the stimuli, in Experiments 1 and 4, we drew the noise for only half of the Fourier coefficients at random and chose the other half as the respective complement to preserve the symmetry in coefficients. Figure 1b displays examples of such stimuli for two different levels of phase noise.

Contrast gradients

To investigate the effect of global changes in luminance contrast independently from other features in the image, we increased or decreased luminance contrast gradually in the

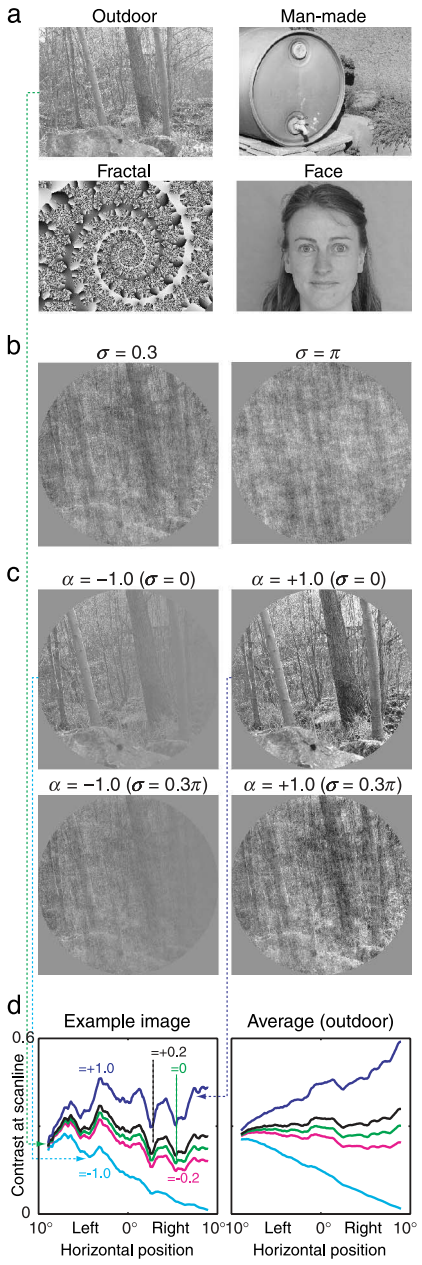

Figure 1. Stimuli. (a) Examples of the four image categories used. All images were grayscale. (b) Phase noise. Upper left stimulus of Panel a at two different levels of phase noise; radius of aperture spans 10°. (c) Contrast gradients. Stimuli with contrast gradients to the right. (d) Contrast along the horizontal midline for left-to-right gradients (left: outdoor image of Panel a; right: average over all outdoor images).

horizontal direction (Figure 1c). The luminance of the modified stimulus $I(x, y)$ was computed from the original $I_0(x, y)$ for left-to-right gradients,

$$I(x,y) = \left(\alpha\frac{(x-1)}{(L-1)}\right)(I_0(x,y) - \langle I_0\rangle) + I_0(x,y), \quad (1)$$

and for right-to-left gradients,

$$I(x,y) = \left(\alpha\frac{(L-x)}{(L-1)}\right)(I_0(x,y) - \langle I_0\rangle) + I_0(x,y), \quad (2)$$

where $1 \le x \le L$ and $1 \le y \le L$ are given in pixels and L is the width of the image (L = 768 in Experiments 2 and 3 and L = 1,024 in Experiments 1 and 4). $\langle I_0\rangle$ denotes the mean over all pixels, and α determines the slope of the gradient. To measure the effect of strong gradients as well as the effect of gradients approximating the size of variations naturally occurring in natural scenes, we used values of α of -1.0 and -0.2 (reduction in contrast as compared with the original; Figure 1c, top), 0 (no gradient), and +0.2 and +1.0 (increase in contrast; Figure 1c, bottom).

Local contrast modifications

In Experiment 4, we used—besides the gradient modified stimuli—images with local contrast modifications (Einhäuser & König, 2003) on one side. These modifications were constructed by using a set of Gaussian masks, G_i, each centered at a location (x_i, y_i):

$$G_i(x,y) = \exp\left(-\frac{\left((x-x_i)^2 + (y-y_i)^2\right)}{\lambda^2}\right). \quad (3)$$

Taking the maximum over G_i resulted in the overall mask:

$$G(x,y) = \max_{i\in\{1,\ldots,P\}}[G_i(x,y)]. \quad (4)$$

In analogy to Equations 1 and 2, each image point of the original pixel intensity $I_0(x, y)$ was then modified to

$$I(x,y) = I_0(x,y) + \alpha G(x,y)(I_0(x,y) - \langle I_0\rangle). \quad (5)$$

In addition to the modification strength, α, we get two more parameters: the number of modifications, P, and their size, λ. In each locally modified image, we used one of three different sizes: λ = 320 (large), λ = 160 (medium), and λ = 80 (small). To approximately match the integrated modification, we chose the number of modified regions (P) to be 1, 5, and 25, respectively. Modifications were restricted to one side of the image (modified side). To avoid excessive overlap with image edges or midline, we restricted the potential range of modification centers. For modifications on the left, we used modifications centers $220 < x_i < 292$ (large modifications), $100 < x_i < 412$ (medium), and $80 < x_i < 432$ (small); analogously, for modifications to the right, we used $732 < x_i < 804$, $612 < x_i < 924$, and $592 < x_i < 944$, respectively. Within these

ranges, the centers of the local modifications were randomly chosen so that no two modifications would be less than a distance λ apart.

Definition of luminance contrast

In line with earlier studies (Reinagel & Zador, 1999) and as a canonic generalization of the two-point contrast, we defined luminance contrast at each point of a stimulus as the standard deviation of luminance in a patch divided by the mean luminance of the stimulus. In the present context, it is important to note that the global gradients as described above do manipulate luminance contrast according to this definition. Figure 1d depicts the luminance contrast measured along a horizontal scan line in an image modified with the different gradient strengths, α. It can be seen that local luminance contrast is dominated by the gradient, on average. Hence, the definition of luminance contrast used in the analysis of local features and the gradients is compatible. We based all analysis on a patch size of 80 × 80 pixels (2.1° × 2.1°), but results were qualitatively similar for a wide range of patch sizes.

Contrast at fixation/choice of baseline

We aim to test whether or not contrast is elevated at fixated locations, as compared with feature values measured at randomly sampled locations (“baseline locations”). There are different possible means of choosing the baseline locations. First, one can sample the image uniformly, that is, measure the feature value at or around each pixel. We will refer to this sampling as “uniform baseline.” Comparing fixated locations to the uniform baseline, however, may be subject to a confound: Assume that the feature under investigation is likely to be elevated at a certain region (say, the center). Assume further that fixations are generally biased toward this region. Then, even if the feature has no effect on fixation, comparison between fixated locations and uniform baseline locations would show elevation of the feature at fixated locations. Hence, we also compute a second baseline, which we sample at all locations fixated by the subject when viewing all other stimuli of the same category and gradient level. Because these locations account for effects of general biases in fixated locations, we will refer to this sampling as “unbiased baseline” (for a thorough discussion of central biases and the appropriate choice of baseline, see also Mannan et al, 1996; Tatler et al., 2005). Here, we report results relative to the unbiased baseline throughout. Results were not qualitatively different for the uniform baseline.

Experimental design

Experiment 1 (stimulus categories)

In this experiment, we tested four different image categories (outdoor scenes, man-made objects, fractals, and faces). Each subject was shown two different, randomly chosen images of each category, at 12 different phase-noise levels, yielding a total of 2 × 4 × 12 = 96 trials per subject. Subjects were instructed to “study the images carefully.” Each trial lasted 6 s and was preceded by a central fixation cue on medium-luminance background. After 50 trials, the calibration of the eye-tracking system was validated and the system was recalibrated if needed. All but one subject performed one session (96 trials). For the lone subject who performed two sessions, only the first session was used for analysis. Including the second session, however, did not qualitatively affect the results.

Experiment 2 (contrast gradients)

In this experiment, 10 outdoor scene images were presented at 10 different levels of additive phase noise, yielding 100 different stimuli. These stimuli were used without gradient ($\alpha = 0$) and with four different gradient levels ($\alpha = -1.0$, $\alpha = -0.2$, $\alpha = +0.2$, and $\alpha = +1.0$) in two directions (from left to right and from right to left). The resulting 900 different stimuli were distributed over nine blocks of 100 trials. Stimuli were balanced such that each of the 100 combinations of picture and phase-noise level appeared exactly with one type of gradient in each block.

Each trial started with a black central fixation cue on a medium-luminance (55 cd/m^2) background that was displayed for 0.5 s. Subsequently, the stimulus was presented for 2.5 s. After the offset of the stimulus, subjects had to indicate by pressing one of two mouse buttons “whether or not this image appears natural, i.e. resembles the image of a real-world scene” (direct quotation from the written instructions given to the subject before the experiment). Following the subject’s response, the next trial started immediately.

Before each block, the calibration of the eye-tracking device was validated and a new calibration was performed if necessary. Between blocks, subjects were allowed to take breaks. Subjects performed three to five blocks in each recording session and needed two or three sessions to complete all of the nine blocks.

Experiment 3 (two-alternative forced-choice experiment)

This experiment used the same images and noise levels as Experiment 2. Ten different versions with different random patterns of phase noise were generated for each of these 100 conditions. The resulting 1,000 stimuli formed the target set for the two-alternative forced-choice (2-AFC) experiment.

In each trial, 2 stimuli were presented in succession: 1 of the 1,000 target stimuli and 1 distracter that shared the same amplitude spectrum but had random phase drawn from a uniform distribution between $-\pi$ and π. Each of the two stimuli was presented for 0.5 s and preceded by 0.5 s of medium-luminance blank screen. Subjects were asked to indicate, “which of the two stimuli looked more natural, i.e. more closely resembled the image of a real-world scene.”

The order of trials was random throughout the experiment. The 1,000 trials were subdivided into 40 blocks of

25 trials. After each block, there was a break of at least about 50 s, which subjects could extend as needed.

Experiment 4 (modification size)

In this experiment, we tested whether the effect of contrast gradients was comparable to the effect of local contrast modifications. We used 10 outdoor images (distinct from those used in Experiments 2 and 3) at the same modification levels used in Experiment 2 ($\alpha = \pm 1.0$ and $\alpha = \pm 0.2$). Besides the gradients, we used three different modification sizes (large: $\lambda = 320$, medium: $\lambda = 160$, and small: $\lambda = 80$). Because we were primarily interested in the effect of modification size, we used only the two extreme phase-noise levels: no noise and random phase. Instructions were identical to Experiment 2. Using 10 images at two noise levels, two directions, four modification strengths, and four modification sizes (gradients and three different λ), in addition to 10 images without contrast modification at two noise levels, yielded 660 trials. These were randomly ordered and subdivided into 11 blocks of 60 trials, after each of which subjects could take a break when needed.

Presentation

Stimuli in all experiments were computed and displayed using Matlab (MathWorks, Natick, MA) and its psychophysics toolbox extension (Brainard, 1997; Pelli, 1997) running on a Windows PC.

- In Experiment 1, stimuli were presented on a 21-in. CRT monitor (Samsung Electronics Co. Ltd., Korea) located 80 cm from the subject.
- In Experiments 2 and 3, stimuli were presented on a 19-in. CRT monitor (Sony, Tokyo, Japan), which was located 85 cm in front of the subject. Maximum luminance ("white") of the presentation screen was 110 cd/m^2, whereas ambient light levels were below 0.01 cd/m^2. The fringes of the stimuli were masked by a circular aperture with radius of 10° visual angle ($L/2 = 384$ pixels) to reduce any effects of screen boundaries. The background outside the aperture had medium luminance (55 cd/m^2). The gamma of the screen was corrected to ensure a linear mapping from pixel values to displayed luminance.
- In Experiment 4, we used a different monitor (Dell Inc., Round Rock, TX), whose maximum luminance was 29 cd/m^2 with otherwise identical settings to Experiments 2 and 3.

In all experiments, subjects' heads were stabilized at constant distance from the screen using a chin rest and a forehead rest.

Data acquisition

Throughout Experiments 1, 2, and 4, we recorded observers' eye positions using a noninvasive infrared eye tracker. Experiment 1 used an Eyelink 2 system (SR Research Ltd., Osgoode, ON, Canada); Experiment 2 used an ISCAN ETL-400 (ISCAN, Burlington, MA, USA), and Experiment 4 used an Eyelink-1000 system. For the two Eyelink systems, we used the manufacturer's software for calibration and validation and for determining periods of fixation. For the ISCAN system, the mapping to screen coordinates was computed using a grid of 25 predefined fixation locations as the bilinear transformation that minimizes the mapping error for these data points. We determined fixations for the ISCAN by using the algorithm developed by Peters et al. (2005; Peters, personal communication). We verified that both algorithms that determine fixations yield comparable results on identical data sets.

Subjects

Fourteen volunteers (20 to 28 years old) from the University of Osnabrück participated in Experiment 1. The same five volunteers from the Caltech Community (19 to 28 years old) participated in Experiments 2 and 3. Five additional volunteers from the Caltech Community (20 to 33 years old) participated in Experiment 4. All subjects had normal or corrected-to-normal vision, were naive as to the purpose of the experiment, and were paid or given course credit for participation. All experiments conformed to the Declaration of Helsinki and to the National and Institutional regulations for experiments with human subjects.

Results

Behavioral data: Are higher order statistics needed for perceiving scenes as natural?

First, we analyze to what extent the subjective perception of a stimulus as natural depends on phase noise (behavioral reports of Experiments 2, 3, and 4). In Experiment 2, subjects performed a yes/no paradigm. Images that did not undergo changes in their phase spectra ($\sigma = 0$) were judged as natural in almost all cases (99.4%, Figure 2a). This value decreased with increasing phase noise in a highly nonlinear fashion, which suggests a categorical rather than a continuous transition. It reached 9.0% for the maximum noise level ($\sigma = \pi$). This general behavior does not depend on the gradient strength ($p = .18$, two-factor ANOVA on the factors of noise level and gradient strength). Some subjects, however, still reach relatively high values for $\sigma = \pi$ (maximum: 26%), which suggests that individuals might employ different criteria. During Experiment 4, when there were only two noise levels (no noise and random phase), observers showed the same pattern of responses: They judged the no-noise stimuli almost always natural (98.4% ± 1.4% natural [$M \pm SD$]) and the random-phase stimuli "non-natural" (2.9% ± 2.4% natural). To obtain a criterion-free

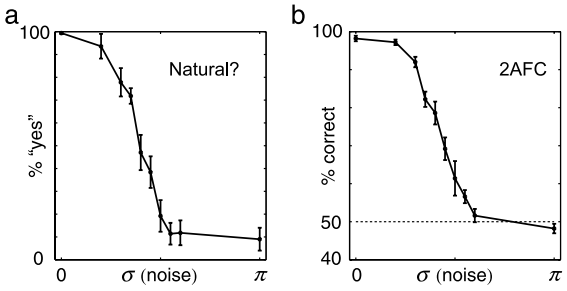

Figure 2. Behavioral data. (a) Percentage of "yes" responses on whether or not an image is natural in Experiment 2 versus phase-noise level. (b) Percentage correct (image with less phase noise is judged more natural) in Experiment 3 versus phase-noise level. In both panels, error bars denote the standard error of the mean over the $n = 5$ subjects.

measurement, we used a 2-AFC paradigm in Experiment 3. The subjects of Experiment 2 had to report which of the following images they found more natural: one with varying σ or one with random phase ($\sigma \rightarrow \infty$). For $\sigma = 0$, subjects almost always (98.2%) correctly judge the image with less phase noise as "more natural." This performance decreases again in a sigmoidal fashion, becoming indistinguishable from chance level for $\sigma \geq 0.6\pi$ (t test: $p = .41$ for $\sigma = 0.6\pi$ and $p = .22$ for $\sigma = \pi$, Figure 2b). These data demonstrate that the subjective perception of an outdoor scene as natural is contingent on higher order stimulus statistics and that the transition between the two categories is sharp.

Does the relation between contrast and fixation depend on higher order statistics?

As first analysis of eye-movement data, we tested to what extent the relation of luminance contrast to fixation depends on higher order statistics (Experiment 1). We tested 12 levels of phase noise and four different stimulus categories: outdoor scenes, fractals, man-made objects, and faces. For stimuli without noise, the mean luminance contrast was elevated at fixated locations relative to baseline in all categories (outdoor scenes: 0.9% ± 0.6%; fractals: 7.2% ± 2.6%; man-made objects: 6.8% ± 5.0%; faces: 2.9% ± 1.5%, relative elevation over unbiased baseline, expressed as $M \pm SEM$ across subjects, Figure 3). For all categories, the mean elevation across noise levels is significantly larger than 0 ($p = .001$, $p = 2 \times 10^{-4}$, $p = .01$, $p = .005$, t tests for the individual categories pooled over noise levels). However, the elevation decreases with increasing phase noise, and its size is significantly anticorrelated to η, the amount of noise ($r = -.95$, $p = 3 \times 10^{-6}$; $r = -.63$, $p = .03$; $r = -.89$, $p = 1 \times 10^{-4}$; $r = -.83$, $p = 9 \times 10^{-4}$ for the four categories). The results of Experiment 1 confirm the elevation of luminance contrast at fixated locations for a variety of complex stimuli. They furthermore demonstrate a dependence of this *correlative* effect on higher order statistics. This is first evidence that a higher order property related to contrast might contribute to the elevation of contrast at fixated locations.

Do contrast gradients bias fixation?

Next, we measured the extent to which large-scale contrast gradients bias observers' eye position (Experiment 2). For each 2.5-s trial, we measured the median horizontal eye position over the whole trial, irrespective of the type of eye movement (fixation, saccade, etc.). For images without any gradient ($\alpha = 0$), the mean eye position is 0.02° ± 0.8° ($M \pm SD$ over subjects) right of the center, which is not a significant bias ($p = .96$, t test). Large positive gradients ($\alpha = +1$) introduce a strong bias toward the side of higher contrast, which is significant for left-to-right gradients (0.86° ± 0.42° to the right, $p = .01$, Figure 4a) and shows the same tendency for right-to-left gradients (0.96° ± 1.07° to the left, $p = .11$). The tendency is preserved for small gradients $\alpha = +0.2$, which bias the eye position 0.25° ± 0.54° and 0.24° ± 0.94° in the direction of the gradient, although the effect is not significant for this gradient strength ($p = .36$ and $p = .60$, respectively). For negative gradients, we observe a similar effect: The bias always goes in the direction of the higher contrast, yielding significant biases of 1.64° ± 1.12° ($p = .03$) and 0.99° ± 0.77° ($p = .04$) for $\alpha = -1.0$ as well as similar nonsignificant tendencies for $\alpha = -0.2$ (0.35° ± 0.73° and 0.21° ± 0.91°, $p = .35$ and $p = .64$, respectively, Figure 4a). Although the bias is not significant for shallow gradients, the gradient biases observers' eye position toward higher contrasts in all cases. To quantify this effect further, we measure the difference between the biases in eye position for opposing gradients for each subject and for each gradient strength α (Figure 4b). Over subjects, this difference is significantly different from 0 for all but one α ($p = .01$, $p = .01$, $p = .12$, and $p = .008$ [t tests] for $\alpha = -1.0$, $\alpha = -0.2$,

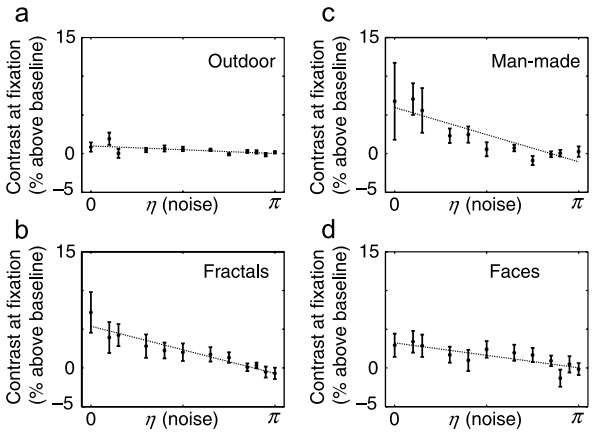

Figure 3. Elevation of contrast at fixations. Luminance contrast at fixations relative to unbiased baseline (see Methods section). (a) Outdoor scenes, (b) fractals, (c) man-made objects, (d) faces. All data are expressed as mean and standard error of the mean over subjects and best linear regression.

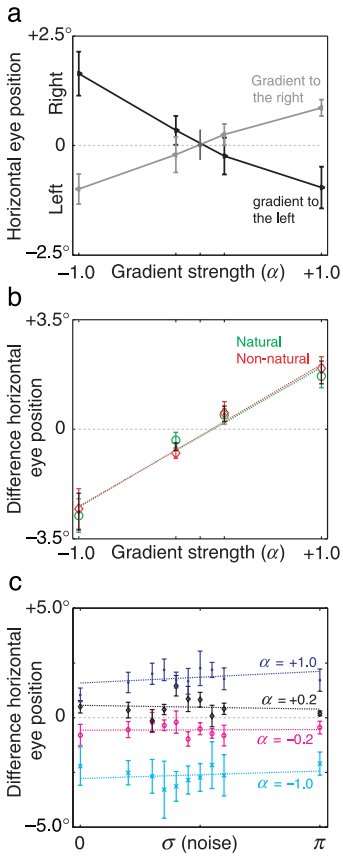

Figure 4. Effect of contrast gradients. (a) Horizontal eye position in trials with gradients from left to right (gray) and gradients from right to left (black) for different gradient strengths α. 0° denotes center of screen. (b) Difference between horizontal eye position in trials with left-to-right gradients and right-to-left gradients for all stimuli (black), stimuli that subjects judged to be natural (green), and stimuli that they judged nonnatural (red). Lines denote corresponding optimal linear fit. (c) Difference between horizontal eye positions between left-to-right and right-to-left gradient trials versus phase-noise level. Different colors denote different gradient strengths. All panels show mean and standard error of the mean over subjects.

α = +0.2, and α = +1.0, respectively) and is highly significantly correlated to α (r = .99, p = .007).

In the Appendix, we analytically derive that a linear relation of fixation probability to contrast is consistent with this linear correlation. In summary, the contrast gradient effectively biases eye position toward regions of higher contrasts, consistent with a linear relation of fixation probability and luminance contrast.

Does the fixation bias induced by contrast gradients depend on higher order statistics?

Next, we investigate the extent to which the biases induced by first-order features depend on higher order statistics. Grouping the eye-tracking data by the response of the subject reveals no difference between natural (Figure 4b, green) and nonnatural images (Figure 4b, red): Neither is there any difference for any α (p = .81, p = .22, p = .86, and p = .66) nor is the correlation between eye-position difference and α affected (natural: r = .98, p = .02; nonnatural: r = .995, p = .005). In addition to the binary grouping in natural and nonnatural images, we also analyze how the effect of contrast gradients depends on the phase-noise level. We do not find any significant correlation between the effect of the contrast gradient on eye position and phase noise for any α (α = −1.0: r = .23, p = .53; α = −0.2: r = .03, p = .94; α = +0.2: r = −.10, p = .78; α = +1.0: r = .39, p = .26; Figure 4c). The same pattern of results is found when performing the analysis for periods of fixations only (data not shown). In conclusion, the effect of the contrast gradient on eye position is significant and linear in the gradient strength α. In addition, it is independent of whether or not the stimulus is judged as natural.

Do local contrast modifications have the same effect as large-scale gradients?

Although we find that contrast reductions repel eye position in this study, earlier studies (Einhäuser & König, 2003; Einhäuser et al., 2006) had demonstrated that strong *local* reductions of contrast have an attractive effect. Although these findings themselves stand undisputed, their interpretation has spurred controversy (Kayser et al., 2006; Parkhurst & Niebur, 2004). Consequently, we tested whether the size of the modifications reconciles these findings (Experiment 4). Consistent with the data of Experiment 2, contrast gradients biased observers to the side of high contrast. This bias had a trend to linear correlation with α (r = .94, p = .07) and was independent of phase noise for all α (t tests: p = .99, p = .99, p = .12, and p = .72 for α = −1, α = –0.2, α = 0.2, and α = 1, respectively). As with the gradients, there is no significant difference between no-noise and random-phase stimuli at any modification level or modification size ($p > .19$ for all t tests for α and λ). For all three sizes, the bias depends on modification level (large: p = .002, medium: p = .001, small: p = .004, ANOVA). However, a significant linear correlation with modification level is observed only for the largest modification size (r = .98, p = .02, Figure 5a) but not for medium (r = .92, p = .08, Figure 5b) or small (r = .80, p = .20, Figure 5c) modification. Furthermore, for the small and medium modifications, the dependence on α is not monotonic, with strong negative modifications (α = −1.0) being less repulsive than moderate ones (α = −0.2). Finally, small, strongly negative modifications are—if anything—attractive ($M \pm SD$: 0.01° ± 0.59° toward modified side). These results are in line with earlier data on local modifications and make it conceivable that strong modifications assume an object-like quality that attracts attention, counteracting the otherwise repulsive effect of low contrast.

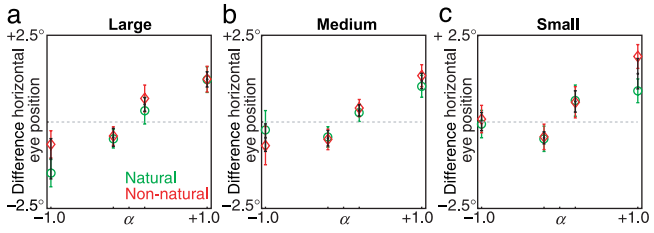

Figure 5. Eye-position biases induced by local modifications of different size. Difference between horizontal eye position in trials with modifications on the right and on the left; black: all stimuli; green: no noise; red: random phase. (a) Large modification (λ = 320); (b) medium modification (λ = 160); (c) small modification (λ = 80).

Discussion

In this study, we investigate how higher order stimulus statistic modulates the effect of contrast on fixation. We demonstrate that

1. The perception of an outdoor scene as natural is contingent on higher order stimulus statistics.
2. For a variety of categories, luminance contrast is elevated at fixated locations. This elevation is anti-correlated to phase noise and, thus, depends on higher order statistics.
3. Global modifications of luminance contrast (gradients) attract attention if contrast is increased and repel attention if contrast is decreased. This effect is independent of higher order statistics.
4. The repulsive effect of decreased contrast vanishes or reverses for local modifications.

In summary, luminance contrast biases attention as predicted by bottom–up models. However, to explain the local effects of contrast in natural scenes in full, one needs to consider correlations to higher order statistics.

The predominant sensory-driven (bottom–up) model of human attention, the saliency map (Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985), is based on difference maps (contrasts) in first-order features, such as luminance. Such saliency maps predict human fixation locations significantly above chance (Itti & Koch, 2000; Parkhurst et al., 2002; Peters et al., 2005; Tatler et al., 2005). However, their predictions are still far from the theoretical optimum for bottom–up models—the mutual prediction of scan paths between different observers (Oliva, Torralba, Castelhano, & Henderson, 2003; Peters et al., 2005; Privitera et al., 2005). In line with the present results, this suggests that incorporating relations to higher order statistics may improve such bottom–up models.

Two approaches exist to incorporate higher order statistics into bottom–up models. First, one may select a different set of features than the classical saliency map, which explicitly or implicitly include higher order effects: Among the popular choices, there are edge density (Mannan et al., 1996); localized edges, corners, and points (Krieger et al., 2000); "texture contrast" (Parkhurst & Niebur, 2004, see below); and localized generic geometrical kernels (Privitera et al., 2005). Alternatively, one may learn the relevant features from scene statistics. Such a learning approach comes at the advantage that it can be specific to an image category (Torralba, 2003; Torralba & Oliva, 2003), which modulates at least effects of some low-level features (Parkhurst et al., 2002) and simple geometric properties like symmetry (Privitera & Stark, 2000). Learning approaches can furthermore readily be extended to incorporate task-specific priors, that is, top–down knowledge (Navalpakkam & Itti, 2005; Oliva et al., 2003; Torralba, 2003). Irrespective of the preferred modeling approach, our present findings highlight the importance of higher order structure for the effect of a low-level feature—luminance contrast.

Several studies have investigated the relation of luminance contrast to human attention in natural stimuli. Most of these studies find that luminance contrast is elevated at fixated locations (Einhäuser & König, 2003; Krieger et al., 2000; Mannan et al., 1997; Reinagel & Zador, 1999; Tatler et al., 2005). The range of effects we observe here is consistent with these results, when taking into account systematic biases of observers and images, as done in this study (see Mannan et al., 1996; Tatler et al., 2005, for a thorough discussion of this issue). Consequently, we confirm earlier studies in describing a small, though significant, elevation of luminance contrast at fixated locations.

In an earlier study (Einhäuser & König, 2003), we had demonstrated that local reductions of luminance contrast attract attention. Consequently, the elevation of luminance contrast at fixations is not a consequence of contrast itself. Instead, the elevation of contrast at fixation is the consequence of their mutual correlation to a higher order property. This interpretation is in line with the present data, in which contrast elevation anticorrelates with noise level; that is, it is at least partly dependent on higher order statistics.

Using a modified saliency-map model, Parkhurst and Niebur (2004) argued that a measure of contrast variation, which they dubbed texture contrast, is elevated when contrast is locally decreased. Assuming texture contrast to be 10-fold more attractive than luminance contrast, their model indeed reproduced some aspects of the Einhäuser and König (2003) data. The fact that contrast gradients and large modifications bias attention to high contrasts, where small negative modifications have no such repulsive effect (Experiment 4), provides an alternative explanation: Strong local negative modifications deviate from the local surrounding. Thereby, they stick out as an odd item, much like isolated features in pop-out (Treisman & Gelade, 1980). In the temporal domain, such local

deviations from global context or expectation, recently formalized as Bayesian "surprise" (Itti & Baldi, 2005), also attract attention. Hence, it is well conceivable that strong negative modifications form a deviation from context and, therefore, attract attention, counteracting the repulsive effect of reduced contrast. This view and the texture-contrast interpretation of Parkhurst and Niebur are not mutually exclusive. On the contrary, texture contrast forms one possible formalization of this concept, which also highlights the relative importance of higher order stimulus statistics for the *local* guidance of overt attention.

Appendix

Linear model for contrast biases

Here, we demonstrate how the assumption that contrast has a linear effect on fixation probability results in the bias induced by gradients to be linear in gradient strength α. The conditional probability to fixate a contrast c is given as

$$p(\text{fix}|\text{contrast} = c) = \kappa c + n, \tag{A1}$$

where n is a normalization constant and κ parameterizes the effect of contrast on fixation. For simplicity, we consider here only left-to-right gradients. Assuming that the gradient of strength α dominates the average contrast position x (Figure 1d), the contrast $c(x)$ is given as

$$c(x) = c_0\left(\alpha\frac{x}{L} + 1\right), \tag{A2}$$

where L is the length of the image and c_0 is the unmodified contrast (Figure 1d). The probability to fixate a certain contrast is

$$p(\text{fix}, c) = p(c)p(\text{fix}|c). \tag{A3}$$

Under Equation A2, the prior for a contrast to be in the image is uniform in $[c_0, (\alpha + 1)c_0]$ for positive α (or in $[(\alpha + 1)c_0, c_0]$ for negative α) and 0 outside:

$$p(c) = \begin{cases} \frac{1}{|\alpha|c_0} & c_0 < c < (\alpha+1)c_0 \\ \frac{1}{|\alpha|c_0} & c_0 > c > (\alpha+1)c_0 \\ 0 & \text{otherwise} \end{cases}. \tag{A4}$$

Assuming all fixations are on the image, we have

$$p(\text{fix}, c) = \frac{1}{|\alpha|c_0}(\kappa c + n), \tag{A5}$$

with p being a probability we get for the constant n, for positive α

$$\begin{aligned} 1 &= \int_{c_0}^{(\alpha+1)c_0} p(\text{fix}, c)\text{d}c \\ &= \frac{1}{\alpha c_0}\left(\left[\frac{\kappa}{2}c^2\right]_{c_0}^{(\alpha+1)c_0} + [nc]_{c_0}^{(\alpha+1)c_0}\right) \\ &= \left(\kappa c_0\left(\frac{\alpha}{2}+1\right) + n\right) \\ \Rightarrow n &= 1 - \kappa c_0\left(\frac{\alpha}{2}+1\right). \end{aligned} \tag{A6}$$

By interchanging the lower and upper integration limits and replacing α by $-\alpha$ in the normalization, we obtain the same result for negative α. Hence, we expect the fixated contrast (again, the notation assumes positive α, but exchanging the integration limits and replacing $|\alpha| = -\alpha$ in the normalization yields the same result for negative α) to be

$$\begin{aligned} \langle c\rangle &= \int_{c_0}^{(\alpha+1)c_0} cp(\text{fix}, c)\text{d}c \\ &= \frac{\kappa}{\alpha c_0}\int_{c_0}^{(\alpha+1)c_0} c^2\text{d}c + \frac{1}{\alpha c_0}\int_{c_0}^{(\alpha+1)c_0} c\text{d}c - \frac{\kappa}{\alpha}\left(\frac{\alpha}{2}+1\right)\int_{c_0}^{(\alpha+1)c_0} c\text{d}c \\ &= \frac{\kappa c_0^2\,\alpha^2}{12} + c_0\left(1+\frac{\alpha}{2}\right). \end{aligned} \tag{A7}$$

Plugging this result into Equation A2, we obtain the mean eye position to be expected for gradient α as

$$\begin{aligned} \langle c\rangle &= c(x_\alpha) \\ &= c_0\left(\alpha\frac{x_\alpha}{L}+1\right) \\ \Rightarrow x_\alpha &= \frac{L}{\alpha}\left(\frac{\langle c\rangle}{c_0} - 1\right) \\ &= \frac{L}{\alpha}\left(\frac{\kappa c_0^2\,\alpha^2}{12c_0} + \frac{c_0}{c_0}\left(1+\frac{\alpha}{2}\right) - 1\right) \\ &= \frac{L\kappa c_0\alpha}{12} + \frac{L}{2}. \end{aligned} \tag{A8}$$

Analogous to the above calculation, for right-to-left gradients, one obtains

$$\Rightarrow x_\alpha = \frac{L}{2} - \frac{L\kappa c_0 \alpha}{12}. \quad \text{(A8')}$$

This implies that a gradient extending over the whole image biases fixation by a fraction

$$\frac{\kappa c_0 \alpha}{12} \quad \text{(A9)}$$

of the image width L, relative to the image center ($L/2$) in the direction of the gradient. This analytical result has two important consequences: If one assumes a linear model for the effect of contrast (or any feature) on fixation probability and the gradient is sufficiently strong (compared with the naturally occurring variation of the feature), one predicts that

1. The induced position bias is linear in the gradient strength,
2. A gradient in one direction of strength α has an equivalent effect to the opposing gradient of strength $-\alpha$.

Both predictions are consistent with our observations for gradients.

Acknowledgments

This work was supported by the Swiss National Science Foundation (W.E., Grant Nos. PBEZ2-107367 and PA00A-111447), NIMH, NGA, NSF, and Caltech's "SURF" program. We thank A. Acik and H.-P. Frey for technical assistance and C. Quigley for editorial assistance.

Commercial relationships: none.
Corresponding author: Wolfgang Einhäuser.
Email: wet@klab.caltech.edu.
Address: CIT 216-76, 1200 E. California Blvd., Pasadena, CA 91125, USA.

References

Betsch, B. Y., Einhäuser, W., Körding, K. P., & König, P. (2004). The world from a cat's perspective—Statistics of natural videos. *Biological Cybernetics, 90,* 41–50. [PubMed]

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

Buswell, G. T. (1935). *How people look at pictures. A study of the psychology of perception in art*. Chicago, IL: The University of Chicago Press.

Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology, 12,* R476–R477. [PubMed] [Article]

Einhäuser, W., & König, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience, 17,* 1089–1097. [PubMed]

Einhäuser, W., Kruse, W., Hoffmann, K. P., & König, P. (2006). Differences of monkey and human overt attention under natural conditions. *Vision Research, 46,* 1194–1209. [PubMed]

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. [PubMed]

Itti, L., & Baldi, P. (2005). A principled approach to detecting surprising events in video. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (vol. 1, pp. 631–637).

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506. [PubMed]

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20,* 1254–1259.

James, W. (1890). *Principles of psychology*. New York: Holt.

Kayser, C., Nielsen, K. J., & Logothetis, N. K. (2006). Fixations in natural scenes: Interaction of image structure and image content. *Vision Research, 46,* 2535–2545. [PubMed]

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4,* 219–227. [PubMed]

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher-order statistics. *Spatial Vision, 13,* 201–214. [PubMed]

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision, 10,* 165–188. [PubMed]

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception, 26,* 1059–1072. [PubMed]

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45,* 205–231. [PubMed]

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top–down control of visual attention in object detection. *IEEE Proceedings of the International Conference on Image Processing, 1,* 253–256.

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

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 107–123. [PubMed]

Parkhurst, D., & Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes. *The European Journal of Neuroscience, 19,* 783–789. [PubMed]

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

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom–up gaze allocation in natural images. *Vision Research, 45,* 2397–2416. [PubMed]

Privitera, C. M., Fujita, T., Chernyak, D., & Stark, L. W. (2005). On the discriminability of hROIs, human visually selected regions-of-interest. *Biological Cybernetics, 93,* 141–152. [PubMed]

Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22,* 970–982.

Rainer, G., Augath, M., Trinath, T., & Logothetis, N. K. (2001). Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Current Biology, 11,* 846–854. [PubMed]

Rainer, G., Lee, H., & Logothetis, N. K. (2004). The effect of learning on the function of monkey extrastriate visual cortex. *PLoS Biology, 2,* E44. [PubMed] [Article]

Reinagel, P., & Zador, A. (1999). Natural scene statistics at the centre of gaze. *Network, 10,* 341–350. [PubMed]

Rizzolatti, G., Riggio, L., Dascola, I., & Umilta, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia, 25,* 31–40. [PubMed]

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Physical Review Letters, 73,* 814–817. [PubMed]

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45,* 643–659. [PubMed]

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention, *Cognitive Psychology, 12,* 97–136. [PubMed]

Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 20,* 1407–1418. [PubMed]

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems, 14,* 391–412. [PubMed]

van der Schaaf, A., & van Hateren, J. H. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research, 36,* 2759–2770. [PubMed]

Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. R. (2006). Phase noise and the classification of natural images. *Vision Research, 46,* 1520–1529. [PubMed]

Yarbus, A. L. (1967). *Eye movements and vision* (B. Haigh, Trans.). New York: Plenum.