Spatiotemporal mechanisms for detecting and identifying image features in human vision

Peter Neri and David J. Heeger

Department of Psychology, Serra Mall 450, Stanford University, Stanford California 94305, USA
Correspondence should be addressed to P.N. (pn@white.stanford.edu)

Published online: 8 July 2002, doi:10.1038/nn886

Our visual system constantly selects salient features in the environment, so that only those features are attended and targeted by further processing efforts to identify them. Models of feature detection hypothesize that salient features are localized based on contrast energy (local variance in intensity) in the visual stimulus.This hypothesis, however, has not been tested directly. We used psychophysical reverse correlation to study how humans detect and identify basic image features (bars and short line segments). Subjects detected a briefly flashed `target bar´ that was embedded in `noise bars´ that randomly changed in intensity over space and time. By studying how the intensity of the noise bars affected performance, we were able to dissociate two processing stages: an early `detection´ stage, whereby only locations of high-contrast energy in the image are selected, followed (after ~100 ms) by an `identification´ stage, whereby image intensity at selected locations is used to determine the identity (whether bright or dark) of the target.

Visual processing, at all stages, involves selection of relevant information. When humans navigate in the environment, the visual system analyzes only a small fraction of the incoming stimulation[1,2] because cortical resources are limited, as is the time window during which one can respond appropriately and effectively to outside stimuli. It is therefore crucial that this initial selection be carried out efficiently. Salient features in the retinal image are those most likely to be important to the animal[3]. For example, a strong variation in luminance giving rise to an edge in the image is likely to be associated with the boundary of an object, and therefore with an important characteristic of the three-dimensional layout of the environment.

To investigate how the human visual system detects and identifies such simple image features, we assessed those aspects of the stimulus that are important in the detection of bright and dark bars briefly flashed on a screen. We used an extension of a technique known as noise-image classification[4,5], a psychophysical variant of noise-based reverse correlation[6]. In the first of two experiments, we asked subjects to detect the presence or absence of a bar embedded in spatiotemporal noise. We found that detection of this simple object involved two stages: an early stage at which subject responses were influenced by variance in noise intensity, followed by a later stage at which performance depended on the polarity of the noise. We then tested the hypothesis that these two stages reflect the two different processing steps of detecting and subsequently identifying the feature[7]. In a second experiment, subjects carried out these two tasks (detection and identification) on every trial[8,9]. We analyzed which properties of the noise influenced performance in each of these two tasks, and found that detection judgments depended on early variance in noise intensity, whereas identification judgements depended on the polarity of the noise after 50–100 ms. These results are consistent with a large body of literature on bottom-up, exogenous attention[10–13], as well as with models of feature detection[14] and models of attentional selection[15–17].

Results

Subjects viewed stimuli consisting of a temporal sequence of nine frames, each containing eleven vertical, spatially adjacent bars. On half the trials, a bright bar (target) appeared in the middle of the central frame in the stimulus (**Fig. 1a**), and subjects reported its presence or absence. Their task was made more difficult by adding a random intensity to all bars in all frames, thus generating a noisy stimulus that varied in both space and time (**Fig. 1b**). The spatiotemporal profile for the noise bars (**Fig. 1c**) varied from trial to trial owing to the random sampling procedure. Some of these variations in the noise distribution had no effect on detection performance, but others were highly influential. After each trial, the computer recorded the intensity of noise bars at each location in each frame. The individual trials were sorted into the four possible outcomes—hit, false alarm, miss or correct rejection. The individual noise distributions were separately averaged (μ) for each of these four classes, and we computed the variance (σ^2) of the noise distributions within each class (**Fig. 1d**).

We then combined the four averages of the noise images, adding the average noise distributions associated with 'yes' responses and subtracting those associated with 'no' responses, to yield a mean kernel: a function that describes how first-order statistics (that is, average) in the noise affect performance[4]. Similarly, we combined the variances of the noise images to yield a variance kernel, which describes the effect of second-order statistics (that is, variance) in the noise on subjects' performance.

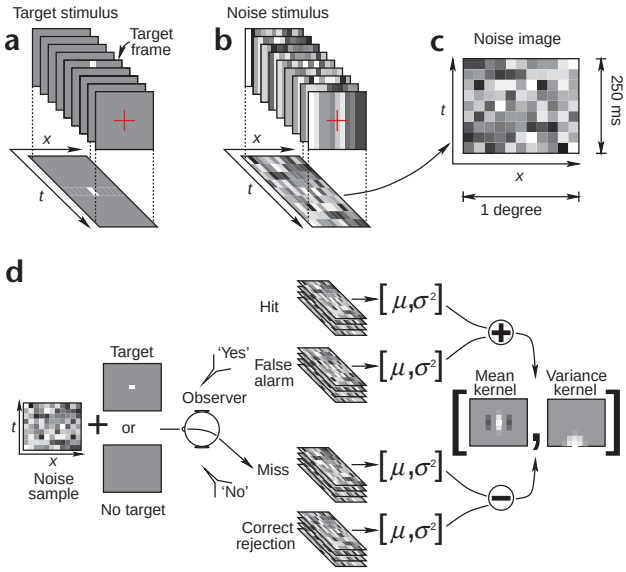

Fig. 1. Stimuli and reverse-correlation technique. The stimulus consisted of a sequence of 9 frames, each divided into 11 vertical bars (**a** and **b**). (**a**) The target was a bright bar in the middle of the central frame in the sequence. A red fixation cross marked the center of the stimulus (target location), and never disappeared. (**b**) Noise consisted of a random luminance modulation of all bars in all frames. (**c**) Each noise image spanned 250 ms (27 ms per frame) in time and 1 degree in space. (**d**) The target was added to the noise on half the trials (left), and subjects were asked to detect its presence. Noise images for each trial were classified according to the response class (hit, false alarm, miss, correct rejection). Mean (μ) and variance(σ^2) kernels were computed from the noise images for each response class, as indicated (Methods).

The noise statistics (mean and variance kernels) indicated whether the particular spatiotemporal distribution of noise intensity on each trial assisted or hindered detection of the target. We found that particular spatiotemporal noise distributions tended to cause subjects to make particular behavioral responses, thus allowing us to infer some properties of the visual detection and identification mechanisms for salient features. First, 'yes' responses (hits and false alarms) were more frequent than 'no' responses when there happened to be an excess of noise intensity close to the spatiotemporal location of the target. This is evident in the mean kernels, which had a positive peak at the spatiotemporal location of the target, and negative flanks across the spatial dimension (**Fig. 2**, top row, first and third panels). Variance kernels also showed a peak, but this peak was mainly confined to a temporal period that preceded the target by ~100 ms (**Fig. 2**, top row, second and fourth panels). In a similar experiment, subjects were asked to detect a bar of the opposite polarity (dark rather than bright), and mean kernels showed a negative peak at the spatiotemporal location of the target (**Fig. 2**, lower plots). Variance kernels were very similar to those seen for detecting a bright bar, again showing an early peak. The temporal offset between the two peaks was statistically significant, for both subjects and for both bright and dark bar targets (dotted ellipses in **Fig. 2**).

The temporal offset between peaks in the mean and variance kernels suggests that the two kernels probably reflect different mechanisms. Because peaks in variance kernels do not carry information about the polarity of the target and occur before peaks in mean kernels, which do carry such information, we reasoned that the two kernels reflect two separate processing steps[7,13]. The first step detects and localizes salient visual features based on the contrast energy in the stimulus, and the second step identifies those features based on the pattern of stimulus intensities (bright or dark bars, in this case).

To test this hypothesis, we performed an experiment that addressed the issues of detection and identification[8,9] directly. In experiment 2, the target bar appeared bright or dark, and either to the left or to the right of fixation. Subjects were asked to determine both the location and the polarity of the target on each trial[9]. Noise distributions were then analyzed separately for the two tasks of detecting and identifying the target (Methods). Detection generated pronounced modulations in the variance kernel, but not in the mean kernel (**Fig. 3a**, upper plots). The opposite was true for identification (**Fig. 3a**, lower plots). The lack of significant modulation in the variance kernel for identification was not due to lack of a sufficiently large data set, because the smaller data set in experiment 1 generated significant modulations. This confirms our hypothesis that these two stages are separately reflected by the two kernels. Specifically, detection judgements (when subjects reported 'left' or 'right') were not differentially dependent on polarity (whether the noise was bright or dark); hence the mean kernel for detection is zero.

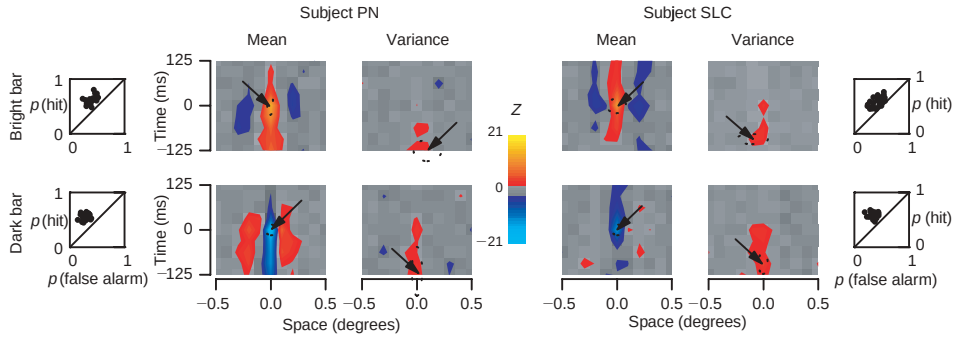

Fig. 2. Mean and variance kernels for two subjects. Insets, each dot reports probability of hit versus probability of false alarm[28] for an individual block of 200 trials. Upper plots, detecting a bright bar (5,000 trials per subject); lower plots, detecting a dark bar (5,000 trials per subject). For each subject, both mean and variance kernels are shown. Color (linearly interpolated data), *Z* scores (–21 to 21) for $|Z| > 2$. Arrows indicate the largest peak in each plot. Dotted ellipses are 95% confidence intervals for the peaks (Methods).

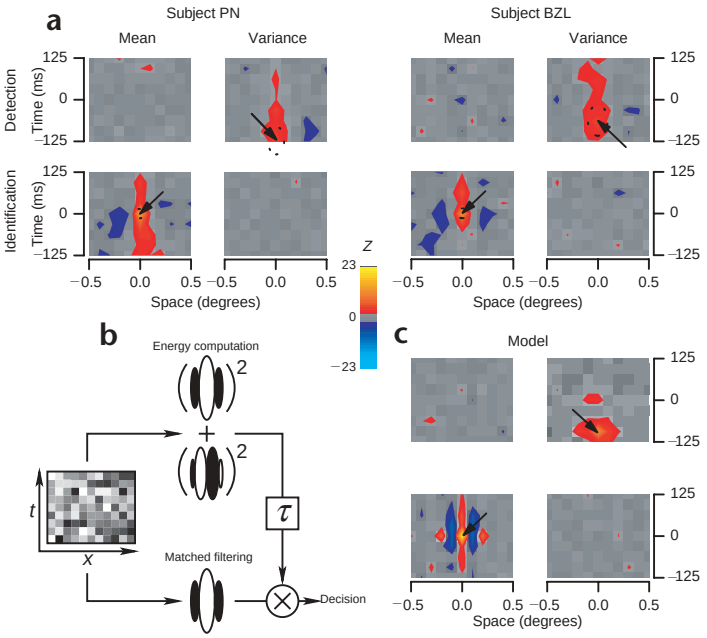

Fig. 3. Experiment 2. (**a**) Mean and variance kernels for two subjects (5,000 trials for each subject) for both detecting (upper plots) and identifying (lower plots) the target bar. Color indicates linearly interpolated data, *Z* scores (–23 to 23) for $|Z| > 2$. Format as in **Fig. 2**. Detection involved significant modulation only in variance kernels, and identification only in mean kernels. (**b**) Model used to simulate the experimental data. Stimuli underwent two stages, contrast energy extraction and linear matched filtering. Contrast energy modulated (after some temporal delay $\tau = 100$ ms) the gain (×) of the matched-filters, and the output from the latter was used to simulate behavioral responses. (**c**) Simulated kernels (compare with experimental kernels in **a**). Percent correct for detection, 64% (PN), 60% (BZL); for identification on correctly detected trials, 69% (PN), 68% (BZL); for identification on mislocalized trials, 50% (PN), 49% (BZL) (consistent with ref. 8).

What matters to the localization judgement is variance early in the stimulus, as demonstrated by the variance kernel. Similarly, the lack of modulation in the variance kernel for identification implies that identification judgements (when subjects reported 'bright' or 'dark', after the target had been localized) are not differentially dependent on variance, but rather on polarity (as shown by the mean kernel). Notably, detection (variance kernel) peaked before identification (mean kernel) for both subjects.

Discussion

The measured variance kernels reflect a non-optimal strategy for our psychophysical task. When the stimulus had a high level of contrast energy, subjects were more inclined to report that the target was present. This tendency, however, led to more errors in detection because contrast energy presented ~100 ms before the target provides no information about the target itself. The ideal strategy would involve monitoring stimulus intensity only at the precise time and location of target presentation. This indicates that our experiments probed built-in mechanisms that are automatically engaged by stimulus presentation, and which are probably useful in many real-life contexts[2,15,16]. These mechanisms are not subject to cognitive strategies and cannot be 'switched off' in conditions in which they are inefficient.

We think our results are best interpreted with relation to automatic, exogenous, bottom-up attentional capture by high-contrast cues[10]. An early burst of contrast energy can be thought of as a high-contrast cue that drives subjects' attention to its location, thus affecting further processing[10,11]. The temporal offset of ~100 ms between detection and identification is consistent with a large body of literature on exogenous cueing[10–13]. Moreover, neurons that are selective for a target among distractors have been shown to manifest this selectivity roughly 100 ms after an initial nonspecific increase in firing for both target and distractors[18]. This 100-ms temporal offset is also consistent with a variety of electrophysiological studies of attentional selection[2,19–22].

To illustrate our interpretation of these psychophysical data, we implemented a model of target detection and identification, which shares some elements with previous models of attention[15–17] and with previous models of feature detection[14]. In our model, detection and localization depend on extracting contrast energy from the stimulus, and identification depends on matched filtering (**Fig. 3b**). The responses of the matched filters are modulated, after a brief delay, by the extracted contrast energy so that detection by the latter leads to subsequent identification (and behavioral decision) by the former. We ran the same experimental procedure with this model, using exactly the same values for all parameters. The model captured most aspects of the psychophysical results (**Fig. 3c**). To generate the temporal offset between peaks in the two kernels, the delay (τ in **Fig. 3b**) was set to 100 ms. The critical features of this model are that (i) the detection mechanism depends largely on stimulus contrast energy, and (ii) detection precedes identification by ~100 ms. The implementation presented here is only intended as a basic starting point, not as a comprehensive or quantitative description of the complex perceptual processing that must be operating in human subjects.

Our results offer insight into the way in which the human visual system achieves fast and efficient processing of basic image features. Because neural resources are limited and processing must be completed quickly to generate appropriate behavioral responses, the visual system cannot carry out detailed processing of all stimuli that reach our eyes[2,16]. Early processing is therefore aimed at detecting and selecting salient features in the

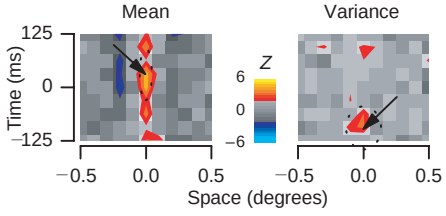

Fig. 4. Mean and variance images for false alarm trials only, from experiment 1, bright bar target. Detection of a dark bar was similar, only with a mean kernel of inverted polarity as in **Fig. 2**. Data was pooled from both subjects (PN and SLC, total of 2,108 false alarms). Same format as in **Figs. 2** and **3**. The variance image reports variance modulations around the baseline noise variance.

image[15–17,23], allowing cortical resources to be subsequently directed only to behaviorally relevant locations[1,2,16]. Our results indicate that this early selection stage involves contrast energy extraction. Contrast energy is a robust indicator of the presence of salient features in the image, although it is a poor indicator of their exact structure[24]. Once salient parts of the image have been selected (possibly leading to eye-movement planning[25–27]), further processing is carried out at those locations to identify the exact nature of the feature (for example, whether it is a bright or a dark line). The neuronal mechanisms underlying these processes, along with their precise computations, are sure to be the subject of further study.

Methods

Task. Stimuli were generated by a VSG graphics card and presented on a CRT monitor (Vision Master 17, Iiyama, Cambridge Research Systems, Rochester, UK) at a viewing distance of 57 cm. Subjects fixated a central red cross against a gray background (35 cd/m^2). This fixation marker was centered on the stimulus (**Fig. 1**) and never disappeared. In the context of experiment 1, this means that there was no spatial uncertainty as to the location of the target. Stimuli consisted of nine frames (27 ms each). For experiment 1 (**Fig. 2**), each frame contained 11 vertical, spatially adjacent bars (0.1° × 1.1° each), centered around fixation. Each bar in each frame was assigned a random luminance value drawn from a discrete uniform distribution spanning ± 4 cd/m^2 around background luminance (**Fig. 1b**), in 0.4 cd/m^2 steps. The target consisted of a 4 cd/m^2 increase (or decrease, when subjects were asked to detect a dark bar) in the central bar of the middle frame in the temporal sequence (**Fig. 1a**), randomly applied on half of the trials. For experiment 2 (**Fig. 3**), each frame contained 22 bars of the same size, 11 to the left and 11 to the right of fixation. Each bar in each frame was assigned a random luminance value over the range ± 3.6 cd/m^2. The target bar consisted of either an increase or a decrease of 4 cd/m^2 in the bar centered on either the 11 bars to the left or those to the right of fixation, again only for the middle frame in the sequence. A target was present on each trial, but randomized for polarity and location.

At the end of each presentation, subjects made a forced-choice decision. For experiment 1 (subjects had knowledge of the polarity of the bar), the decision was between presence or absence of the target bar. For experiment 2, subjects reported whether the target bar appeared to the left or to the right of fixation and whether it was bright or dark. Subjects indicated their decision with a button press which then triggered the next stimulus presentation after an interval randomly varying between 200 and 500 ms. No feedback was provided. Blocks consisted of 200 trials. Every 25 trials, a target bar without noise was presented so as to remind the subject of its appearance and to help to maintain vigilance[28]. All experiments were subject to Human Subjects Protocol Approval by the National Institutes of Health, and all subjects gave informed written consent.

Reverse-correlation analysis. Our technique is an adaptation and extension of a previous technique[4] that was developed for a Vernier acuity task. For the experiments involving detection only (**Fig. 2**), the spatiotemporal intensity distribution of the noise for each presentation was stored and sorted according to which response class had occurred (hit, false alarm, miss or correct rejection). We then computed both average (μ) and variance (σ^2) for the noise distributions (**Fig. 1c**) corresponding to each response class, and combined them as follows:

$$\text{Mean kernel}\ (x.t) = \mu_{[1,1]}(x.t) + \mu_{[0,1]}(x.t) - \mu_{[1,0]}(x.t) - \mu_{[0,0]}(x.t)$$
$$\text{Variance kernel}\ (x.t) = \sigma^2_{[1,1]}(x.t) + \sigma^2_{[0,1]}(x.t) - \sigma^2_{1,0]}(x.t) - \sigma^2_{[0,0]}(x.t)$$

where $\mu_{[s,r]}$ and $\sigma^2_{[s,r]}$ are the mean and variance images across noise images of trial type $[s,r]$, where s refers to the target (0 for absent, 1 for present), and r to the response from the subject.

For the experiments involving simultaneous detection and identification (**Fig. 3**), each noise image was divided in two halves: one to the left and one to the right of fixation. For computing kernels for detection (**Fig. 3a**, upper plots), these halves were separately classified by taking subjects' responses as double statements of the sort "Target present on the left, target absent on the right" or "Target absent on the left, target present on the right" (equivalent to ref. 29). This made it possible for each half on each trial to be sorted into one of the four response classes. For example, if the target bar was presented on the left and the subject reported its presence on the right, the left-half noise image was classified as a miss, and the right-half noise image as a false alarm. For identification (**Fig. 3a**, lower plots), kernels were computed using the equations above (with s and $r = 1$ for bright target, $= 0$ for dark target), only on those trials and for those half images corresponding to correct detection. For example, when a bright bar was correctly detected on the left but misidentified by the subject to be dark, the left-half noise image was classified as a miss.

Statistical analysis. We used a statistical bootstrapping procedure[30] to estimate the reliability of the resulting kernels. That is, we estimated the standard deviation of each kernel separately for each point in space and time. These standard deviations were then used to compute the Z scores in **Figs. 2–4**. Bootstrapping was also used to derive a probability function for the location of the peak in each kernel. We then fitted to this function an elliptic Gaussian and plotted its contour at 95% of its volume.

Modeling. Psychophysical data were simulated using a spatial array of energy extractors and matched filters (**Fig. 3b**). Matched-filter responses were computed by convolving the stimulus images, generated exactly as for the psychophysical experiments, with a linear filter (even-symmetric in space and monophasic in time). Energy responses were modeled as the squared sum of a quadrature pair of linear filters[31,32], f_{even} and f_{odd}. Separate computations were done for each (left and right) half of each stimulus, yielding a pair of spatiotemporal maps of energy and matched-filter responses, one for each half of each stimulus in each simulated trial of the experiment. The spatiotemporal map of energy responses was shifted temporally by $\tau = 100$ ms, and was then used to modulate the gain of the matched-filter responses (via multiplication). The resulting gain-adjusted, matched-filter responses were then multiplied by a 'discriminability' map D that weighed each point in space and time according to the information it could possibly provide about the target. The discriminability map was computed to make optimal use of the matched-filter responses, based on the average and standard deviation of the responses across trials[28]. The final output for the left half of the stimulus can be expressed as:

$$\text{Output}_{\text{left}} = \iint\left[\left(f_{\text{even}}(x,t-\tau) * I_{\text{left}}(x,t)\right)^2 + \left(f_{\text{odd}}(x,t-\tau) * I_{\text{left}}(x,t)\right)^2\right] \cdot \left[f_{\text{even}}(x,t) * I_{\text{left}}(x,t)\right] D(x,t)\,\mathrm{d}x\,\mathrm{d}t$$

$$D(x,t) = \frac{E\left(f_{\text{even}}(x,t) * I^1_{\text{left}}(x,t)\right) - E\left(f_{\text{even}}(x,t) * I^{-1}_{\text{left}}(x,t)\right)}{\sigma\left(f_{\text{even}}(x,t) * I^{-1}_{\text{left}}(x,t)\right)}$$

$f_{\text{even}}(x,t) = G(\sigma_t, t) \cdot G(\sigma_x, x) \cdot \cos(\omega_x, x)$, $f_{\text{odd}}(x,t) = G(\sigma_t, t) \cdot G(\sigma_x, x) \cdot \sin(\omega_x x)$

$G(\sigma, x)$ is a Gaussian function ($\sigma_t = 40$ ms, $\sigma_x = 1.5$ arcmin, $\omega_x = 4.5$ cycles/°). The subscripts 'left' and 'right' refer to the left- or right-half stimulus image; when the superscript is specified, it refers to presence of a bright (1) or dark (–1) target bar. The output for the stimulus right-half image was computed by substituting I_{right} for I_{left}, obtaining $\text{Output}_{\text{right}}$. The following decisional rule was then applied: location of the target is on the side yielding the largest absolute output ($|\text{Output}_{\text{left}}|$ versus $|\text{Output}_{\text{right}}|$), and identity of the target is bright if this value is greater than 0, dark otherwise.

Alternative models. As an alternative possibility, we considered models based on a single mechanism with temporal and/or spatial uncertainty[33] to explain the results in experiment 1. False alarms comprise the class that is the most informative and for which uncertainty effects (expected to be

constant across trials of the same type) are most pronounced[34]. The experimental mean and variance images for the false alarm class averaged across subjects show that two mechanisms were still clearly segregated by the two images (**Fig. 4**). This is inconsistent with a single-mechanism hypothesis, and thus confirms the main result in **Fig. 2**.

Acknowledgments

We thank H. Barlow, C. Blakemore, B. Cumming, A. Parker, D. Ringach and B. Zenger-Landolt for useful discussions. P.N. was supported by the McDonnell-Pew Program in Cognitive Neuroscience (while at Oxford) and the Wellcome Trust (while at Stanford). D.J.H. was supported by a National Eye Institute grant.

Competing interests statement

The authors declare that they have no competing financial interests.

RECEIVED 28 MARCH; ACCEPTED 12 JUNE 2002

1. James, W. *Principles of Psychology* (Henry Holt, New York, 1890).
2. Kastner, S. & Ungerleider, L. G. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* **23**, 315–341 (2000).
3. Marr, D. *Vision* (Freeman, New York, 1982).
4. Ahumada, A. J. Jr. Perceptual classification images from Vernier acuity masked by noise. *Perception* **26**, 18 (1996).
5. Ringach, D. L. Tuning of orientation detectors in human vision. *Vision Res.* **38**, 963–972 (1998).
6. Eckstein, M. P. & Ahumada, A. J. Jr. Classification images: a tool to analyze visual strategies. *J. Vision* **2**, 1 (2002).
7. Sagi, D. & Julesz, B. "Where" and "what" in vision. *Science* **228**, 1217–1219 (1985).
8. Tolhurst, D. J. & Dealy, R. S. The detection and identification of lines and edges. *Vision Res.* **15**, 1367–1372 (1975).
9. Thomas, J. P., Gille, J. & Barker, R. A. Simultaneous detection and identification: theory and data. *J. Opt. Soc. Am. A* **73**, 751–758 (1982).
10. Posner, M. I. & Cohen, Y. in *Attention and Performance X: Control of Language Processes* (eds. Bouma, H. & Bowhuis, D. G.) 531–556 (Erlbaum, Hillsdale, NJ, 1984).
11. Nakayama, K. & Mackeben, M. Sustained and transient components of focal visual attention. *Vision Res.* **29**, 1631–1647 (1989).
12. Ziebell, O. & Nothdurft, H.-C. Cueing and pop-out. *Vision Res.* **39**, 2113–2125 (1999).
13. Nothdurft, H.-C. Attention shifts to salient targets. *Vision Res.* (in press).
14. Morrone, M. C. & Burr, D. C. Feature detection in human vision: a phase-dependent energy model. *Proc. R. Soc. Lond. B Biol. Sci.* **235**, 221–245 (1988).
15. Itti, L. & Koch, C. A saliency-based mechanism for overt and covert shifts of visual attention. *Vision Res.* **40**, 1489–1506 (2000).
16. Itti, L. & Koch, C. Computational modeling of visual attention. *Nat. Rev. Neurosci.* **2**, 194–203 (2001).
17. Li, Z. A saliency map in primary visual cortex. *Trends Cogn. Sci.* **6**, 9–16 (2002).
18. Constantinidis, C. & Steinmetz, M. A. Neuronal responses in area 7a to multiple-stimulus displays: I. Neurons encode the location of the salient stimulus. *Cereb. Cortex* **11**, 581–591 (2001).
19. Supèr, H., Spekreijse, H. & Lamme, A. V. Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). *Nat. Neurosci.* **4**, 304–310 (2001).
20. Spitzer, H., Desimone, R. & Moran, J. Increased attention enhances both behavioral and neuronal performance. *Science* **240**, 338–340 (1988).
21. Roelfsema, P. R., Lamme, V. A. F. & Spekreijse, H. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* **395**, 376–381 (1998).
22. Martínez, A. *et al.* Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision Res.* **41**, 1437–1457 (2001).
23. Gottlieb, J. P., Kusunoki, M. & Goldberg, M. E. The representation of visual salience in monkey parietal cortex. *Nature* **391**, 481–484 (1998).
24. Piotrowski, L. N. & Campbell, F. W. A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception* **11**, 337–346 (1982).
25. Kustov, A. A. & Robinson, D.L. Shared neural control of attentional shifts and eye movements. *Nature* **384**, 74–77 (1996).
26. Corbetta, M. *et al.* A common network of functional areas for attention and eye movements. *Neuron* **21**, 761–773 (1998).
27. Parkhurst, D., Law, K. & Niebur, E. Modeling the role of salience in the allocation of overt visual attention. *Vision Res.* **42**, 107–123 (2002).
28. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
29. Abbey, C. K. & Eckstein, M. P. Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. *J. Vision* **2**, 66–78 (2002).
30. Efron, B. & Tibshirani, R. *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
31. Adelson, E. H. & Bergen, J. R.. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2**, 284–299 (1985).
32. Pollen, D. A. & Ronner, S. F. Visual cortical neurons as localized spatial frequency filters. *IEEE Trans. Sys. Cybern.* **13**, 907–916 (1983).
33. Pelli, D. G. Uncertainty explains many aspects of visual contrast detection and discrimination. *J. Opt. Soc. Am. A* **2**, 1508–1530 (1985).
34. Ahumada, A. J. Jr. Detection of tones masked by noise: a comparison of human subjects with digital-computer-simulated energy detectors of varying bandwidths. Thesis (Technical Report No. 29), UCLA (1967).