European Journal of Neuroscience, Vol. 17, pp. 1089–1097, 2003

Does luminance-contrast contribute to a saliency map for overt visual attention?

Wolfgang Einhäuser and Peter König
Institute of Neuroinformatics (University and ETH Zürich), Winterthurer Str. 190, 8057 Zürich, Switzerland

Keywords: eye movements, hierarchy, human, top-down, visual system

Abstract

In natural environments, humans select a subset of visual stimuli by directing their gaze to locations attended. In previous studies it has been found that at fixation points luminance-contrast is higher than average. This led to the hypothesis that luminance-contrast makes a major contribution to a saliency map of visual overt attention, consistent with a computation of stimulus saliency in early visual cortical areas. We re-evaluate this hypothesis by using natural and modified natural images to uncover the causal effects of luminance-contrast to human overt visual attention and: (i) we confirm that when viewing natural images, contrasts are elevated at fixation points. This, however, only holds for low spatial frequencies and in a limited temporal window after stimulus onset; (ii) however, despite this correlation between overt attention and luminance-contrast, moderate modifications of contrast in natural images do not measurably affect the selection of fixation points. Furthermore, strong local reductions of luminance-contrast do not repel but attract fixation; (iii) neither contrast nor contrast modification is correlated to fixation duration; and (iv), even the moderate contrast modifications used fall into the physiologically relevant range, and subjects are well able to detect them in a forced choice paradigm. In summary, no causal contribution of luminance-contrast to a saliency map of human overt attention is detectable. In conjunction with recent results on the relation of contrast sensitivity of neuronal activity to the level in the visual cortical hierarchy, the present study provides evidence that, for natural scenes, saliency is computed not early but late during processing.

Introduction

In analysing complex natural scenes, humans direct their attention to a small subset of the input (James, 1890). In the visual domain, attending to a part of the visual field is often associated with actively shifting one's gaze in that direction. This mechanism is commonly referred to as 'overt attention'. Such eye movements are influenced by the task and past experience of the observer and the stimulus properties (Yarbus, 1967). Most theories on the selection of the focus of attention, which address the latter, 'bottom-up'-driven, part rely on the concept of a 'saliency map' (Koch & Ullman, 1985): The input image is analysed locally with respect to various stimulus properties such as luminance, orientated contours and colour. Retinotopic gradients in the resulting maps (contrasts) are summed up, and the location to be attended is selected by a subsequent winner take all process. In short, regions with high contrasts (e.g. luminance-contrast, colour-contrast, orientation-contrast) attract attention.

As physiological substrates for encoding saliency, besides the idea of distributed emergence from locally competitive interactions (Corchs & Deco, 2002), various individual brain regions have been proposed: Recent studies, which found increased luminance-contrast at the centre of gaze when viewing natural images (Reinagel & Zador, 1999; Krieger *et al*., 2000), suggest that luminance-contrast is a major contributor to the saliency map for visual attention (Parkhurst *et al*., 2002). As neuronal responses in early visual areas are highly sensitive to luminance-contrast, these findings are consistent with the encoding of saliency in these early areas (Lee *et al*., 2002; Li, 2002). By contrast, studies using lesions in brain areas, as well as electrophysiological studies that use specific artificial stimuli, found saliency to be represented in different, nonexclusively visual, brain regions. The pulvinar (Posner & Petersen, 1990; Robinson & Petersen, 1992), the superior colliculus (Posner & Petersen, 1990; Kustov & Robinson, 1996; Horwitz & Newsome, 1999; McPeek & Keller, 2002), the frontal eye field (Thompson *et al*., 1997) and the lateral intraparietal area (Gottlieb *et al*., 1998) have been associated with the encoding of saliency maps. This evidence suggests that the computation of a saliency map is not strictly performed in early visual areas. Furthermore, as luminance-contrast sensitivity decreases in the course of the hierarchy of the visual system (Avidan *et al*., 2002), these findings are in conflict with the above observation of a correlation between overt attention and luminance-contrast. Therefore, the question has to be reconsidered: whether luminance-contrast indeed causally contributes to the selection of fixation points, or just happens to be correlated to attention-attracting higher-order properties of natural scenes. This distinction is decisive for the question on the underlying neural substrate. The present study addresses this issue by using natural visual stimuli with locally modified luminance-contrast while keeping the other stimulus parameters constant. We investigate whether these modifications influence the direction of gaze. In a separate paradigm it is also tested whether the used modifications fall within a perceptually relevant range.

Materials and methods

Subjects

For this experiment five volunteers (two female, three male, age between 24 and 41 years) with uncorrected normal vision were used,

Correspondence: Dr W. Einhäuser, as above.
E-mail: weinhaeu@ini.phys.ethz.ch

Received 11 October 2002, revised 13 December 2002, accepted 20 December 2002

doi:10.1046/j.1460-9568.2003.02508.x

four of which (K.C., P.B., R.S., T.F.) were naïve to the purpose of the experiment and none of whom had previously been exposed to the stimuli used. All experiments were undertaken with the understanding and written consent of each subject. Experiments conformed with the institutional and national guidelines for experiments with human subjects and with the Declaration of Helsinki.

Stimuli

All stimuli used in this study are based on natural photographs of local outdoor environment. They depict scenes of open area or in the forest and no man-made objects are visible on these images. Subjectively, they resemble stills from previously recorded natural stimuli by a camera mounted on the head of a cat (Kayser *et al*., 2003), but are taken with a high quality digital camera (3.3Mega pixel colour mosaic CCD, Nikon Coolpix 995, Tokyo, Japan) for increased spatial resolution. They are then downsampled to a resolution of 1024×768 and converted to 8-bit greyscale using the standard MatLab (Mathworks, Natick, MA, USA) function rgb2gray.

As a measure of contrast we follow the definition of Reinagel & Zador (1999): contrast equals the standard deviation of the luminance within a square image region divided by the mean intensity of the image. This definition canonically extends the two-point contrast and is numerically robust. Unless otherwise stated, the length of the square region was chosen to be 80 pixels. For some aspects of later analysis, images were low-pass filtered. This was carried out by applying a two-dimensional Fourier transform to the image, multiplying it with a Gaussian kernel centred at the DC-component and transforming the result back to image space. The width (HWHM) of the kernel will be referred to as cut-off-frequency (f_{cut}). As the filtering suppresses the contribution of high-frequency components the total contrast is reduced. Note that low-pass filtering was performed for analysis only, and in all cases stimuli were presented at full resolution.

In order to modify contrast within a stimulus without introducing intensity boundaries, we applied the following procedure: Five points (x_i,y_i) were randomly chosen in the image, excluding points closer than 160 pixels to the image boundary. A two-dimensional Gaussian:

$$G_i(x,y) = \exp\{-[(x-x_i)^2 + (y-y_i)^2]/\lambda^2\}$$

with $\lambda = 80$ pixel was centered at each of these points. If any of the points were closer to each other than 160 pixels (i.e. 2λ) different random points were selected. Taking the maximum over G_i resulted in the mask:

$$G(x,y) = \max[G_i(x,y)],\ i \in (1,\ldots,5)$$

At each image point the original pixel intensity I_0(x,y) was then modified to:

$$I(x,y) = I_0(x,y) + \alpha G(x,y) * [I_0(x,y) - \langle I_0 \rangle]$$

where $\langle I_0 \rangle$ denotes the mean of I_0 over the image, and α is the peak contrast modification level. If $I(x,y)$ exceeded the 8-bit range of possible pixel values, the result was cropped to the maximum, respectively, minimum possible value. By this procedure local contrast around the points (x_i, y_i) was increased or decreased, while minimally affecting overall intensity and avoiding the introduction of artificial 'object' boundaries. For the experiments described below contrast modification levels from -0.6 to $+1.0$ were used (Fig. 1A).

Stimuli were generated on a Macintosh G4/800 computer (Apple, Cupertino, CA, USA) using MatLab (Mathworks, Natick, MA, USA) including the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). The stimuli were presented at 1024×768 pixel resolution and 120 Hz refresh rate on a 19-inch computer screen (Hitachi CM 772, Tokyo, Japan), which was located at a distance of 57 cm from the subject.

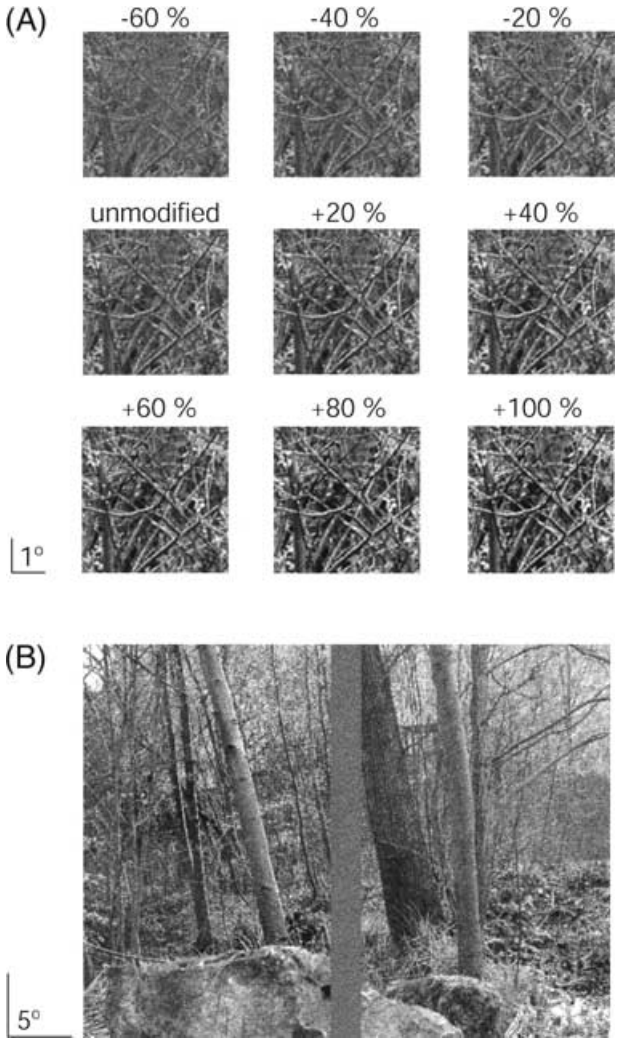


FIG. 1. Stimuli. (A) Illustration of contrast modification. A patch of a natural image subjected to different peak modification levels is shown; the modifying Gaussians G_i are centred at the patch centre, the length of the patch corresponds to two standard deviations of G_i. Note that the application of different modification levels to the same image part is for illustration only. For the experiments, the five locations of modification were chosen randomly and independently for each stimulus-presentation and within one stimulus-presentation all five modifications share the same peak modification level. (B) Stimulus example of the modification detection paradigm. The whole image is subjected to the same type of modification (i.e. five locations are contrast-modified) as in the eye tracking experiment; one half is then replaced by the unmodified image and modified and unmodified part are separated by an overlaid bar. In the example there are three modifications (+100%) visible in the left part of the image (above the bifurcation of the black tree, at the uppermost corner of the rock, at the birch tree at about one-third of the image below the top).

Recording eye position

Eye position was recorded using a commercial oculometer (Dr Bouis, Karlsruhe, Germany; Bach *et al*., 1983). This provides two voltage outputs for horizontal and vertical eye position, respectively, which

were recorded by a CIO-DAS16/JR/CTR-card (Computer Boards, Mansfield, MA, USA), installed into a i80486DX4 PC (Image Microsystems, McLean, VA, USA) at a sampling rate of 1 kHz. From these voltage data, absolute eye position was computed using the following calibration protocol: before each block of stimuli subjects were instructed to subsequently fix points which appeared on an equally spaced 7×7 grid in a 600×600 wide pixel region around the screen centre for 2 s each. The coordinate-transform from oculometer output-voltages to image location chosen for the following stimuli was the bilinear transform which minimizes the sum of squared errors for these 7×7 fixation points. Additionally each stimulus was preceded by a 2-s fixation period, in which the subject was instructed to fixate the centre of the screen. In the fixation as well as the calibration period, fixation points were presented as low intensity (black) crosses on a medium intensity level (grey) background.

The computers used for stimulus presentation and eye-position recording were synchronized through a digital optical signal, which was located in the lower right corner of the presentation screen and covered to be invisible to the subject. In order to minimize head movements subjects were fixed by a forehead rest, a chin rest and a bite-bar made of thermoplastic impression material (Kerr).

Eye tracking paradigm

In order to avoid any task-specific bias in the eye movements, the sole instruction to the subjects regarding the natural stimuli was to 'study the images carefully'.

The experiment consisted of four sessions. The first session for all subjects was composed of three blocks, in each of which each of eight different natural images was presented once for 8 s without any modification and in random order. In each of the sessions 2–4, each image was subjected to nine different peak contrast modification levels ($-60\%, -40\%, \ldots, +100\%$). Using the same eight 'basis' images as in the first session, this yields a total of 72 stimuli per session. These were presented for 8 s in random order and split into six blocks of 12 stimuli. For one of the subjects (P.K.) only peak modification levels from -40% to $+60\%$ were used instead, yielding four blocks of 12 stimuli per session. In all other aspects the protocol was as with the other subjects as described above. Between sessions subjects were allowed to leave the recording set-up.

In all sessions each block was preceded and succeeded by a calibration period and each stimulus by a fixation period. Blocks were excluded from analysis if the root mean square error of the preceding calibration period was larger than 40 pixel ($\lambda/2$), single stimuli were excluded if the error of the preceding fixation period exceeded 40 pixels. We deliberately chose these conservative criteria in order not to miss any potential effects of contrast modification. After applying these criteria, eye traces from a total of 504 (K.C. 153 of 240; T.F. 136/240; R.S. 135/240; P.K. 58/168; P.B. 22/240) stimulus presentations were used for analysis.

Analysis of eye tracking data

To avoid boundary effects and to exploit the most linear range of the oculometer, data analysis was restricted to the part of the recorded eye-traces that fell within the central 600×600 pixel-wide region.

To analyse eye tracking data with respect to the selection of fixation points, different approaches can be taken: First, all recorded data can be analysed without a classification of the underlying eye movement. This avoids introducing arbitrary parameters defining the different types of eye movements. However, a small contamination by image regions grazed during a saccade is included. In a second measure saccades are therefore defined by a velocity threshold, and the gaze directions during a saccade are excluded from data analysis. Both measures implicitly weigh the fixations with their duration. As a third measure, the fixation points as such, i.e. without temporal weighing, are analysed. In the present study we report results of all three types of analysis. Because in our data sample saccades occupy only 12% of the continuously sampled data, we do not observe a qualitative change of results in any case. In all instances of a quantitative deviation between the first two types of analysis both results are given. The third type of analysis additionally addresses the relation of the duration of fixation as a function of contrast, which is interesting in its own right. Therefore, we dedicate a separate section to this analysis.

The distribution of contrasts was computed along the measured eye traces/fixation points of each unmodified image (referred to as 'actual' condition). This result was then compared with the contrast-distribution computed on the same image along all eye traces/fixation points of the same subject obtained on all unmodified images ('control' condition for unmodified images). This procedure excludes those differences between 'actual' and 'control' distributions, which would result from a general bias in eye traces or a consistent nonuniformity of contrast-localization over all images. Whether or not there was a significant bias towards higher (or lower) mean contrast in the actual compared to the control condition was assessed using a two-sided sign test.

In the case of the contrast modified images, for each stimulus presentation the distribution of contrast-modification was computed along the eye trace that was recorded on that stimulus ('actual' condition). Then an 'average' eye trace was generated by concatenating all eye traces/fixation points of the same subject obtained from all presentations of the same basis image. Computing the distribution of contrast modification along the according (i.e. from same subject and basis image) 'average' eye trace for each modified stimulus yields the 'control' condition for that stimulus. It corresponds to the distribution that would be expected if the contrast modification had no influence on fixation.

Separated according to peak contrast modification level, actual and control distributions were then totalled over subjects and basis images. Normalization to unit integral yielded an actual and a control probability density function (PDF) for each peak modification level. The control PDF provides an unbiased prediction of the fixation distribution if the contrast modification had no influence on fixation. That means that if fixation does not depend on contrast modification then control and actual PDF are identical. To statistically measure the similarity of actual and control distribution without any *a priori* assumption a two-sided Kolmogorov–Smirnov (KS) test was performed. As first step the cumulative density functions (CDF) were computed for both conditions at each peak modification level. These CDFs correspond to the integrals over the PDFs. The eye position sampling rate of 1 kHz has no particular rationale regarding the velocity of change of gaze-direction and it was much faster than average fixation time (472 ms). In the first two types of analysis subsequent data-points were therefore not independent. Therefore using the recorded sample length in computing the KS-test *P*-value would overestimate the actual sample size and thus invalidate the *P*-value. To correct for this effect, an effective sample size was calculated as complete recorded sample size divided by average fixation time. Average fixation time was computed as the half-width (FWHM) of the temporal autocorrelation function of the considered eye trace. This effective sample size was used for computing the KS-test *P*-value from the maximum CDF differences.

For the analyses that only use periods of fixation we applied the following procedure to separate periods of fixation from the rest of the data: Any period not interleaved by eye movements faster than a

threshold velocity (v_{thresh}) and lasting at least a minimum duration of T_{min}, is regarded as a single fixation. Gaze location during a fixation is defined as the mean eye position during this period. Unless otherwise stated we used $v_{thresh} = 50\,^{\circ}$/s and $T_{min} = 10$ ms.

Modification detection paradigm

In order to measure the degree to which contrast modification is detectable by subjects, the following two-alterative forced choice protocol was performed in a separate session: modified stimuli were generated as in the eye tracking paradigm, i.e. contrast modifications of identical peak modification level were applied in five randomly chosen locations across the image. Then one randomly chosen half of the image was replaced by its unmodified version. To avoid potential boundary cues between modified locations in one part and their unmodified counterpart, the midline of the complete stimulus was covered by a 60 pixel-wide grey bar (Fig. 1B). Additionally, 50% of the resulting stimuli were mirror-reversed on the vertical midline. Subjects were instructed to indicate by a button press which part of the stimulus was contrast-modified. The stimulus was shown until the subject made its decision and then followed by a 0.5-s blank. At no time were templates of unmodified images presented. Therefore, subjects had to rely on general 'knowledge' of how natural images look to detect the contrast manipulations. The extreme modifications of -60% and $+100\%$ were described as foggy areas and very high contrast structures. Subjective visibility of the moderate contrast modifications was low, and subjects reported that they were often guessing which side contained the modification. However, no feedback was given to the subjects on the correctness of their decision. Five instances of each of the 16 image versions (eight images + eight reversed) and each of the nine contrast modifications were presented in random order, yielding a total of 720 stimuli. For each stimulus presentation, correctness of response as well as reaction time were recorded. To account for intersubject differences in total reaction time, reaction times were normalized to unit standard deviation within each subject.

For this experiment four subjects (three male, one female, age between 23 and 41 years) with normal or corrected to normal vision were used. Apart from the fact that the bite-bar was not used in this protocol, the equipment for stimulus presentation and localization of subjects were identical to the eye tracking task.

Results

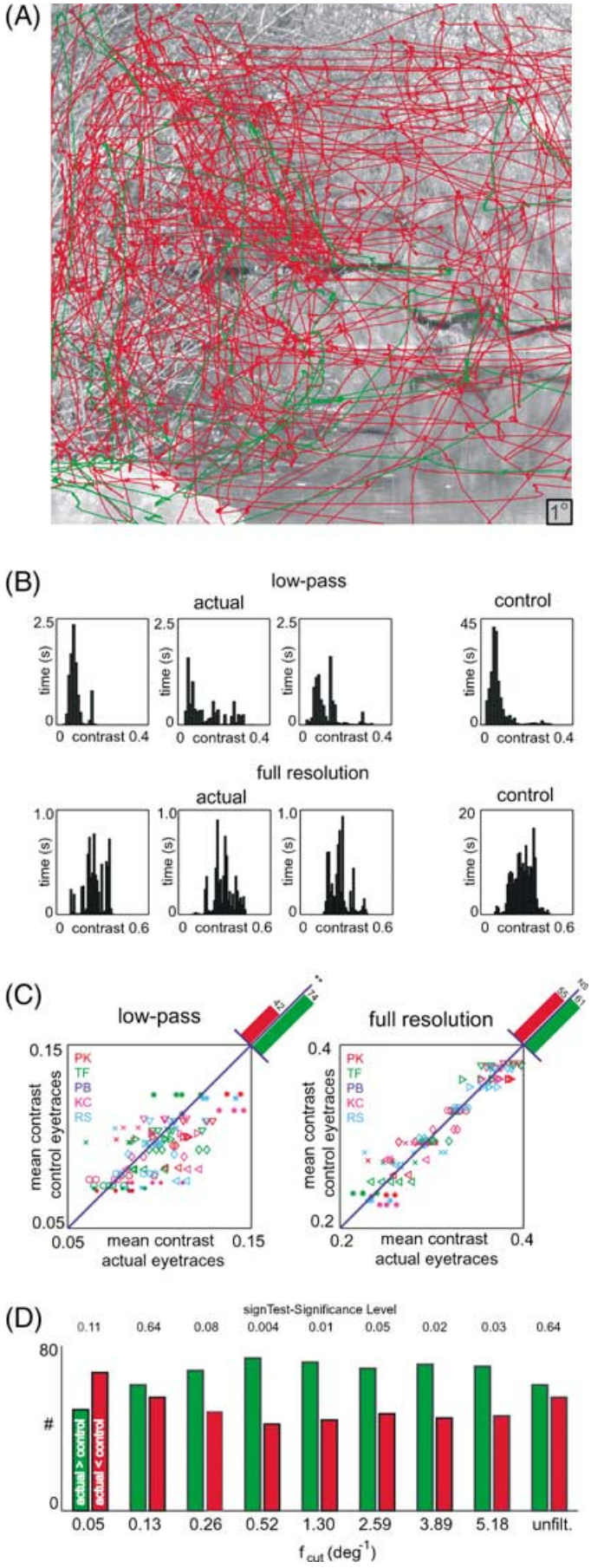
Correlation of contrast and fixation

First we analysed whether contrast correlates with fixation probability in unmodified natural images. For each unmodified image eye traces obtained on the same image ('actual' condition; green traces in Fig. 2A) are compared with eye traces obtained from different images but applied on the same image ('control' condition; red traces in Fig. 2A). If contrast is computed at the resolution used for stimulus presentation, only one out of the three actual traces has higher mean contrast than the control in the example (average relative difference -3%; Fig. 2B, bottom row). On the other hand, if contrast is computed on a low-passed version ($f_{cut} = 0.52$ cyc/$^{\circ}$), contrast is higher for all actual traces compared to the control (average relative difference $+30\%$; Fig. 2B, top row). Note, that low-pass filtering was carried out for analysis only; all stimulus presentations were performed at full resolution. Performing this analysis for all images and all subjects yields a highly significant ($P < 0.01$) difference between 'actual' and 'control' condition if contrast is defined on the low-pass ($f_{cut} = 0.52$ cyc/$^{\circ}$) filtered image (Fig. 2C; left panel) and no significant difference on the unfiltered images ($P > 0.6$; Fig. 2C, right panel). Analysing this frequency dependence further, reveals that the difference between actual and control condition is maximal for cut-off frequencies around 0.5 cyc/$^{\circ}$ and remains significant over a wide range of cut-off frequencies. Nevertheless, very high as well as very low spatial frequency contrast has no significant relation to fixation behaviour (Fig. 2D).

In order to address the influence of different spatial scales, the same analysis for the same cut-off frequencies is performed with 40×40 and 160×160 pixel-wide patches instead of the default 80×80. Although in these cases significant differences ($P < 0.05$) are only observed at the cut-off frequencies 0.26 cyc/$^{\circ}$ (at 40×40 only for including saccades; and at 160×160), 0.52 cyc/$^{\circ}$ (for both 40×40 and 160×160) and at 1.3 cyc/$^{\circ}$ (at 160×160), the qualitative dependence of the difference on cut-off frequency is similar. Furthermore, for all patch sizes the maximum influence of contrast on fixation probability is found for the cut-off frequency of 0.52 cyc/$^{\circ}$. This shows that a fixation preference for higher contrast is only present for contrasts below certain spatial frequencies, whereas high frequency contrast has no relation to the selection of fixation points.

Reinagel & Zador (1999) report that the effects of contrast on fixation probability are strongest in the first 4 s after stimulus onset. Indeed, analysing the time course of contrast at fixation points, a significant effect is observed at low-spatial frequencies during the first 4 s (with $f_{cut} \leq 1.30$ cyc/$^{\circ}$, $P = 0.01$ including saccades and $P = 0.02$ excluding saccades), whereas in the interval from 4 s to 8 s, contrast and fixation are not significantly related (with $f_{cut} = 1.30$ cyc/$^{\circ}$, $P = 1.00$ including saccades and $P = 0.93$ excluding saccades). Investigating this effect further at finer temporal resolution, however, shows that the main contribution to the effect does not result from the period directly following stimulus onset (for $f_{cut} = 1.30$ cyc/$^{\circ}$ including saccades, the 0–2 s $P = 0.40$ and 2–4 s $P = 0.01$; for $f_{cut} = 1.30$ cyc/$^{\circ}$ excluding saccades, the 0–2 s $P = 0.51$ and 2–4 s $P = 0.02$). Despite some variation in the effect's time-dependence across cut-off frequencies, for all 0.52 cyc/$^{\circ} \leq f_{cut} < 5.18$ cyc/$^{\circ}$, the major contribution stems from the 2–4 s time interval. As for the complete dataset, no significant effect can be found in any of the mentioned intervals using the unfiltered images ($P > 0.11$ for all 2 and 4 s intervals). Thus the highest influence of low-frequency contrast on attention is found about 2–4 s after stimulus onset and high frequency contrast has no significant influence on attention in any time interval.

Using each basis-image 30 times for each subject (six times unmodified and 24 modified versions) also allows investigation of the effects of stimulus novelty. We performed the same analysis as above separated for session 1 (unmodified only) and sessions 2–4 (unmodified within modified). The general pattern, that significant influence on fixation probability is found for low-pass filtered images (at $f_{cut} = 0.52$ cyc/$^{\circ}$: session 1 $P = 0.02$ and sessions 2–4 $P = 0.02$ including saccades; all sessions $P = 0.01$ excluding saccades) but not for unfiltered stimuli (session 1 $P = 0.33$ and session 2–4 $P = 0.56$ including saccades; session 1 $P = 0.47$ and session 2–4 $P = 0.56$, excluding saccades), is preserved for either session-type. Thus, previous presentation of the same basis-image has no impact on the direction of attention in our experimental paradigm. Therefore the obtained results neither depend on how often a subject has been exposed to a basis stimulus nor whether the unmodified stimuli are shown within a series of modified stimuli. Supported by the fact that none of the subjects could confidently name the number of used basis images when asked after the experiment and the subjective richness of the stimuli, this is strong evidence that effects of stimulus familiarity because of previous exposures to the same basis image does not influence the direction of overt attention.



Modified images

As a next step we addressed whether the observed correlation is a causal consequence of high contrast attracting attention. This is carried out by locally modifying luminance-contrast in the images. The spatial frequency of these modifications is given by the inverse of the standard deviation of the Gaussian masks G_i (i.e. $\lambda/\sqrt{2}$) to 0.47 cyc/°. The distribution of contrast modification along the eye traces are computed for each contrast-modified stimulus and totalled over subjects and basis images but separated according to peak contrast modification level (Fig. 3A, top row). For each peak modification level, this 'actual' distribution is compared to the according 'control' distribution (see methods), which represents the prediction if the modifications have no influence on fixation (Fig. 3A, bottom row). Thus the similarity of 'actual' to 'control' distribution measures the influence of contrast-modification on the selection of fixation points. This similarity is assessed statistically by the KS-test.

The data obtained are used to test two predictions made by any model that proposes luminance-contrast as a contributor to the saliency map of visual attention. First, these models predict that enhanced contrast (positive modification) would lead to increased fixation probability whereas reduced contrast (negative modification) would lead to decreased fixation probability. Second, as the smallest used modifications already influence the local image contrast profoundly, they would also have significant an impact on attention if it is founded on a luminance-contrast based saliency map. Contrary to the latter prediction, the KS-test yields significant differences between actual and control only for −60% ($P < 0.001$; maximal CDF-difference 0.21 including saccades, 0.22 excluding saccades), −40% ($P < 0.01$; maximal CDF-difference 0.07) and +100% ($P < 0.01$; maximal CDF-difference 0.08). In the range from −20% to +60% the maximal CDF difference remains about constant (between 0.021 and 0.027 including saccades, Fig. 3B, solid line; between 0.022 and 0.029 excluding saccades) and no significant difference between actual and control is found ($P > 0.62$ for all modification levels; Fig. 3C), which is also valid for +80% modification ($P > 0.11$; CDF-difference 0.05). Therefore, only contrast modifications above +80% and below −20% are

FIG. 2. Non-modified images. (A) Eye traces of one subject (K.C.) overlaid over the 600 × 600 pixel-wide (23 × 23°) wide central region of an image, which is used for analysis. Green traces were obtained when presenting the displayed image ('actual'), red traces belong to different nonmodified images ('control'). The grey rectangle in the lower right indicates 1 degree of visual angle. (B) Contrast histograms for eye traces of panel A. All 'actual' eye traces are shown separately, 'control' eye-traces on the rightmost panels. Contrast is defined as standard deviation of a 80 × 80 pixel-wide patch divided by image mean intensity on the full-resolution image (bottom row) or on a low-pass filtered image (top row, cut-off frequency: $f_{cut} = 0.02$ per pixel and 0.52 per degree), respectively. Note, that the generally lower numerical values, for both actual and control contrast, after low-pass filtering follow from the definition of contrast as standard deviation. It decreases as high frequency contributions are suppressed by the low-pass filter. (C) Mean contrast on eye traces over all images and subjects for 'actual' vs. 'control' condition (including saccades). Points below the diagonal imply fixation preference for spots of higher contrast. Left panel shows contrast computed on low-pass ($f_{cut} = 0.52$ per degree) filtered image; right panel shows contrast computed on unfiltered image. Same symbols denote same basis image, same colour denotes same subject. The examples used in panels A and B are identified by the magenta leftward pointing triangles. The histograms on the top right of each plot indicate the number of data-points above (red) and below (green) the diagonal, respectively. Points below the diagonal imply an increased fixation probability at higher contrasts. (D) Number of eye traces showing increased fixation probability for spots of higher (green)/lower (red) contrast in dependence of low-pass cut-off-frequency (f_{cut}). The numbers on top of each graph indicate the sign-test significance level. Data in figures include saccades; exclusion of those yields similar significance levels: 0.08, 0.52, 0.05, 0.002, 0.01, 0.08, 0.03, 0.03 and 0.64 for the respective cut-off frequencies.

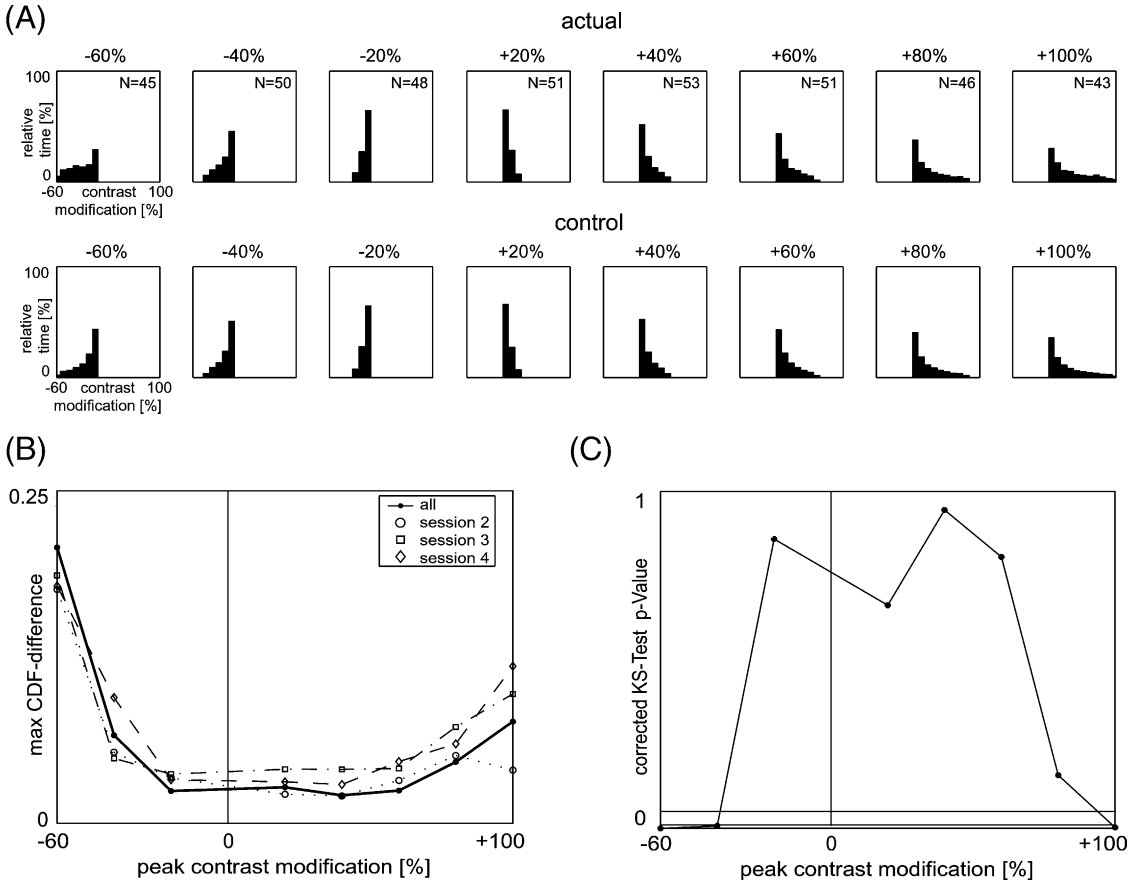


FIG. 3. Modified images. (A) Fixation probability in dependence of contrast modification for different peak contrast modifications (as indicated on top of each plot) for actual (top row) and control condition (bottom row). Histograms of fixation at different peak contrast-modification levels are added over subjects and stimuli and normalized to unit integral. The *N* inset in each top-row panel indicates the number of stimuli used for the according histogram. In control conditions eye traces from all stimuli with non-zero peak modification levels were used ($N_{\text{total}} = 387$) and overlaid over all stimuli at the peak modification level indicated at the top of each histogram. The control condition corresponds to the predicted fixation probability distribution if contrast modification has no impact on fixation probability. (B) Maximum absolute difference between cumulative density functions of actual and control conditions for each peak contrast modification level. Solid line indicates all sessions including modified stimuli (session 2–4); dashed lines indicate individual sessions. The graph is computed including saccades in the analysis. As the maximum relative difference of the plotted values between including and excluding saccades is 7%, the figure applies equally to both types of analysis. (C) *P*-values derived from the CDF difference in panel B using the KS-test corrected for over-sampling (see methods). Low *P*-values indicate significantly different distributions.

found to have a significant influence on attention. Even more decisive for the contribution of luminance-contrast to the saliency map is the result on the first prediction: In all cases in which there is a significant influence of contrast modification on the eye traces, the region of modification always has a higher likelihood to be attended, irrespective of the sign of the modification. This means that locally reduced contrast increases fixation probability at least as strongly as enhanced contrast. This fact, taken together with the large range (-20% to $+80\%$) in which no significant influence of contrast modification is observed, leads to the conclusion that luminance-contrast by itself does not contribute to a saliency map for overt visual attention.

In analogy to the analysis of viewing unmodified images, we here investigate the time course of the effect of contrast modifications on the choice of fixation points. Separating the analysis according to the different experimental sessions shows no systematic effect of the session number on the difference between actual and control condition as represented by the CDFs (Fig. 3B, dotted lines). Thus, as for the unmodified images, the direction of overt attention is not influenced by the memory on previous exposure to the same basis image. Separating the analysis into 2 s and 4 s time-intervals (0–2 s, 2–4 s, 4–6 s, 6–8 s, 0–4 s and 4–8 s) of stimulus presentation yields qualitatively similar results as for the total 8 s; for the -60% peak modification level all intervals show a significantly increased fixation probability for decreased contrast ($P < 10^{-5}$ for all intervals). For the $+100\%$ peak modification level significant differences are only observed for the second half of stimulus presentation ($P > 0.36$ for 0 s to 2 s, 2 s to 4 s and 0 s to 4 s; $P < 0.02$ for all other intervals). Although decreased contrast attracts attention already in the early phase of looking at a stimulus, strongly increased contrast attracts attention only in the later phase of the stimulus presentation. Taken together with moderate contrast enhancements not attracting attention, this finding further supports that there is no causal contribution of contrast to overt attention.

Unweighted analysis of fixation points

Making use of the high sampling frequency of the eye-tracker and using all of the sampled data, the analysis above does not need to distinguish different types of eye movement patterns (saccades, fixations, etc.) and thus can avoid the necessity to select several arbitrary parameters defining these patterns. Although the results are not affected by excluding saccades, the analysis above implicitly weighs fixation probability with fixation duration. Therefore we set out to

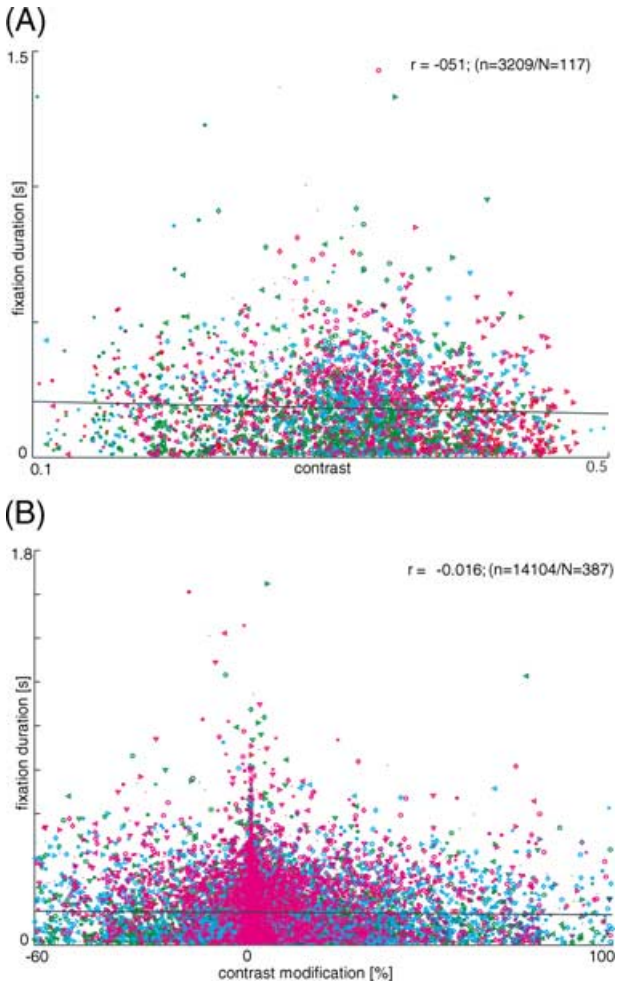


FIG. 4. Fixation duration. (A) Dependence of fixation duration on contrast in unmodified images. Colours and symbols identify subjects and images as in Fig. 2; *N* indicates the number of valid images for each subject, *n* the number of fixations. (B) Dependence of fixation duration on local modification level for modified images; colours and parameters as in panel A.

control for correlation of the latter to contrast and to contrast modifications. Over all unmodified images and all subjects, we found no correlation between contrast and fixation duration ($|r| \leq 0.051$ for all cut-off frequencies; $r = -0.012$ for $f_{cut} = 0.52$ cyc/°; $r = -0.051$ for unfiltered images; Fig. 4A). This holds also true for individual subjects ($|r| < 0.495$ for all subjects and frequencies) and individual images ($|r| < 0.138$ for all images and frequencies), nor do we find any correlation of contrast modification to fixation duration for the modified stimuli ($r = -0.016$; $|r| < 0.112$ for individual subjects, $|r| < 0.076$ for individual basis images, Fig. 4B). In both cases (modified and unmodified images) the results do not depend on the parameters v_{thresh} and T_{min} ($|r| < 0.060$; for all combinations of $v_{thresh} = \{25, 50, 100\}$ °/s and $T_{min} = \{5, 10, 20\}$ ms for unmodified; and $|r| < 0.017$ for the same parameters in the modified case). These controls show that neither contrast nor contrast modification is related to the duration of fixation, and an implicit weighing with fixation duration thus cannot confound the analysis. To address this issue more directly, we repeated the analysis for modified and unmodified images using only the mean eye-position at fixation periods (at $v_{thresh} = 50$ °/s and $T_{min} = 10$ ms) instead of the complete traces. This provides a measure of fixation position which is not weighted by fixation duration. In the unmodified case we find the same qualitative results of the sign-test, although the 'optimal' cut-off frequency is slightly shifted ($P < 0.05$ for $f_{cut} = 1.3$ cyc/°; $P < 0.01$ for $2.6\,\text{cyc/}^\circ \leq^\circ f_{cut}{}^\circ \leq^\circ 5.2\,\text{cyc/}^\circ$; $P > 0.2$ for unfiltered stimuli). In the modified case no significant ($P > 0.05$, KS-test) differences between actual and control condition are found between −20% and +60% peak modification levels, whereas differences at −60%, −40% and +100% are highly significant ($P < 0.001$, KS-test). Most notably, at the +60% peak modification level actual and control conditions are still highly similar ($P = 0.50$, KS-test).

To assess the actual influence of the performed modifications, their impact is compared to the natural contrast distribution in our sample of natural images. On average the analysed 600 × 600 central region of the unmodified images has a mean contrast of 0.30 ± 0.05 (mean ± SD over all images). The average standard deviation of contrast in this region within one image is 0.05 ± 0.02. The average within image standard deviation thus matches the minimally used peak modification level of ±20%, indicating that even the minimal contrast modifications used are in the physiologically relevant range and have a profound influence on local image contrast. By contrast, they do not introduce dramatic global changes to the contrast distribution, as the ±20% peak modification introduce 4% change in the mean contrast over the whole image. Furthermore, the contrast variation across images is similar to the within image variation. This suggests that the basis images are similar to each other regarding their contrast properties, which justifies averaging over images when analysing the data from contrast-modified stimuli.

Modification detection task

The fact, that moderate contrast modifications do not attract attention, but strong modifications of either sign do, raises the question whether the moderate modifications are in a relevant range, i.e. available to the subject on the system level. This issue is addressed in a two-alternative forced choice detection paradigm in which the side containing contrast modifications has to be detected (see methods). These measurements reveal that already a ±20% modification, as all considered non-zero modification levels, can be detected above chance level. Furthermore, it shows that −40% modification, which significantly attracts the gaze, has a similar detection probability ($78\% \pm 6\%$ correct) as +80% modification, which is the maximum positive modification that does not significantly attract the gaze ($81\% \pm 9\%$ correct, Fig. 5A). The analysis of reaction times reveals a similar pattern as the detection probability. The normalized reaction times are slightly smaller for ±20% modification than at 0% and about identical (2.9 ± 1.1 vs. 3.1 ± 1.2) at −40% and +80% (Fig. 5B). As non-gaze-attracting

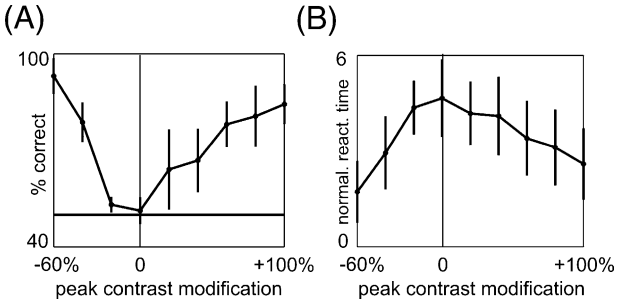


FIG. 5. Contrast modification detection. (A) Percentage correct decisions in the two-alternative forced choice paradigm in dependence of peak contrast-modification level (mean over four subjects). The line at 50% indicates chance level; error bars the standard deviation over subjects. (B) Reaction time over four subjects, normalized to yield unit standard deviation over all peak modification levels within each subject. Error bars denote standard deviation over subjects.

positive modifications show similar detection and reaction-time patterns as gaze-attracting negative modifications, the increased fixation probability at decreased contrast is unlikely to be based solely on a better detectability of the negative modification. This finding does not exclude, however, that the better detectability partly contributes to the increased fixation probability at highly decreased contrasts, as this unnatural modification might consciously attract the attention of the subjects. As the main finding, the modification detection experiment shows that the result, that gaze is not attracted by increased contrast, cannot be explained as consequence of a lack of detectability, i.e availability on the system level.

Discussion

In this study we show that, despite the observed correlation between contrast and fixation probability, high contrast by itself does not causally attract overt attention in natural scenes. As revealed by the detection paradigm, this is true, although the contrast information is available on the system level. In natural scenes luminance-contrast therefore does not contribute measurably to the saliency map for human overt visual attention.

The idea of saliency maps originated in the search for mechanisms that can quickly select possibly interesting locations in complex scenes (Koch & Ullman, 1985; Desimone & Duncan, 1995). Purely bottom-up driven saliency maps dramatically reduce the demand of computational resources and have been proven to perform efficiently even in very complex natural scenes (Itti & Koch, 2000). The relevance of the saliency map idea therefore often derives from its human-like performance even with such difficult stimuli. On the other hand, for the question on the neuronal substrate of saliency maps, the comparison has to be extended from the system level towards the individual processing stages.

Previous studies address the correlation between contrast and fixation probability in natural scenes (Reinagel & Zador, 1999; Krieger *et al*., 2000). Over a certain spatial frequency and temporal range our results confirm their finding that luminance-contrast in (unmodified) natural photographs is higher at attended locations than along random eye traces. The fact that both studies find this effect without low-pass filtering already on their intrinsic – however, arbitrary – spatial scale could be attributed partly to technical differences: the lower resolution of their images, the choice of images and to the patch size the contrast was computed in. Neither of the cited studies, however, has performed an explicit analysis of the influence of spatial scale. The reduced contribution of high frequencies to saccade direction found in the present study seems reasonable, as saccade preparation has to be based on information from higher retinal eccentricities and can thus rely on only reduced spatial resolution. Parkhurst *et al*. (2002) explicitly model the contributions of different low-level stimulus properties (luminance, colour and orientation) to a saliency map. The predictions of their model match their results on human subjects excellently. For natural scenes they find that, among the proposed stimulus properties, luminance-contrast is most salient. Like the studies above, it addresses only the correlation of the low-level features to attention, but does not investigate to what degree they causally determine attention. Thus, it remains open whether the observed predominance of luminance-contrast actually derives from the low-level property itself or from higher-level properties that happen to be correlated to both contrast and attention. This distinction between correlation and causality, however, is decisive to draw conclusions on the underlying neuronal substrate and is therefore dealt with in the present study.

In order to circumvent the problem of confounding higher order properties of complex stimuli with properties of the selection process, most studies on attention rely on more simplified stimuli. Typically stimuli consist of several identical objects, of which one is changed in at least one property (Treisman & Gelade, 1980). This change implies a locally increased feature-contrast and leads the object to 'pop-out' from the background (Nothdurft, 1993). This is also true in the case of increased luminance-contrast. Because it is not obvious whether results obtained on artificial pop-out stimuli can directly be transferred to more natural conditions, it does not, however, contradict the present findings. Indeed, for the particular question on the relative contributions of different stimulus properties to a saliency map, the type of image profoundly influences the results (Parkhurst *et al*., 2002). In the present study natural stimuli are used and modifications are applied such that they are globally within the range of natural variations. Therefore we consider our experiments on natural stimuli well suited to address the contribution of a particular stimulus property under realistic conditions.

As the image type is crucial, care also needs to be taken that the stimuli closely resemble a 'natural' environment. For the present study we chose photographs that are subjectively similar to those obtained by a camera mounted to the head of a freely behaving cat (Kayser *et al*., 2003), apart from the fact that the resolution of the latter had been lower. In particular, unlike for example artistic photographs, street scenes or pictures of urban environment, the present stimuli rarely imply an obvious segregation into clearly defined verbally describable objects. This, along with the fact that the general pattern of contrast (in)dependence does neither change in the course of the experiment nor during viewing an individual stimulus, we are confident that the selected stimuli do not induce an implicit task to the subject. As by the instruction to 'study the images carefully' no explicit task is provided either, the possible task dependent (top-down) influences are reduced as much as possible. Nevertheless, we find that the most direct bottom-up explanation of a saliency map largely based on luminance-contrast is not consistent with our data. Therefore, even if neither an explicit nor an obvious implicit task is involved, top-down processes seem to be dominant for the direction of human visual attention.

Various brain regions have been found to encode stimulus saliency (Posner & Petersen, 1990; Robinson & Petersen, 1992; Kustov & Robinson, 1996; Thompson *et al*., 1997; Gottlieb *et al*., 1998; Horwitz & Newsome, 1999; McPeek & Keller, 2002). All of the suggested areas, higher cortical regions, like the frontal eye field or the lateral intraparietal area, as well as subcortical structures, like pulvinar or the superior colliculus, are strongly interconnected to visual areas. Recent studies measuring (Lee *et al*., 2002) and modelling (Li, 2002) the contribution of early visual areas to saliency map mechanisms also build on pop-out stimuli. Nevertheless, despite the amount of data where saliency is represented, the origin of this representation under natural conditions remains unresolved. Exploiting the different sensitivities to luminance contrast of different cortical areas, however, our results can shed some light on the latter issue; neuronal responses in the early visual system are strongly sensitive to luminance-contrast as tuning curves to most low-level features scale with luminance-contrast (Tolhurst *et al*., 1981; Carandini & Heeger, 1994). While proceeding through the visual hierarchy, on the other hand, responses become more and more invariant to contrast (Avidan *et al*., 2002). Our finding that saliency does not causally depend on contrast thus yields the conclusion that, for natural scenes, the saliency map for overt visual attention does not originate in early cortical visual areas.

Acknowledgements

This work was supported by Honda R&D Europe (Germany) and the Swiss National Science Foundation (SNF–grant No: 31–61415.01).

Abbreviations

CDF, cumulative density function; KS, Kolmogorov–Smirnov (test); PDF, probability density function.

References

Avidan, G., Harel, M., Hendler, T., Ben-Bashat, D., Zohary, E. & Malach, R. (2002) Contrast sensitivity in human visual areas and its relationship to object recognition. *J. Neurophysiol.*, **87**, 3102–3116.

Bach, M., Bouis, D. & Fischer, B. (1983) An accurate and linear infrared oculometer. *J. Neurosci. Meth.*, **9**, 9–14.

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

Carandini, M. & Heeger, D.J. (1994) Summation and division by neurons in primate visual cortex. *Science*, **264**, 1333–1336.

Corchs, S. & Deco, G. (2002) Large-scale neural model for visual attention: integration of experimental single-cell and fMRI data. *Cereb. Cortex*, **12**, 339–348.

Desimone, R. & Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, **18**, 193–222.

Gottlieb, J.P., Kusunoki, M. & Goldberg, M.E. (1998) The representation of visual salience in monkey parietal cortex. *Nature*, **391**, 481–484.

Horwitz, G.D. & Newsome, W.T. (1999) Separate signals for target selection and movement specification in the superior colliculus. *Science*, **284**, 1158–1161.

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

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

Kayser, C., Einhäuser, W. & König, P. (2003) Temporal correlations of orientations in natural scenes. *Neurocomputing*, in press.

Koch, C. & Ullman, S. (1985) Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.*, **4**, 219–227.

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. *Spat. Vis.*, **13**, 201–214.

Kustov, A.A. & Robinson, D.L. (1996) Shared neural control of attentional shifts and eye movements. *Nature*, **384**, 74–77.

Lee, T.S., Yang, C.F., Romero, R.D. & Mumford, D. (2002) Neural activity in early visual cortex reflects behavioural experience and higher-order perceptual saliency. *Nature Neurosci.*, **5**, 589–597.

Li, Z. (2002) A saliency map in primary visual cortex. *Trends Cogn. Sci.*, **6**, 9–16.

McPeek, R.M. & Keller, E.L. (2002) Superior colliculus activity related to concurrent processing of saccade goals in a visual search task. *J. Neurophysiol.*, **87**, 1805–1815.

Nothdurft, H.C. (1993) The role of features in preattentive vision: comparison of orientation, motion and color cues. *Vision Res.*, **33**, 1937–1958.

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

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

Posner, M.I. & Petersen, S.E. (1990) The attention system of the human brain. *Annu. Rev. Neurosci.*, **13**, 25–42.

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

Robinson, D.L. & Petersen, S.E. (1992) The pulvinar and visual salience. *Trends Neurosci.*, **15**, 127–132.

Thompson, K.G., Bichot, N.P. & Schall, J.D. (1997) Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *J. Neurophysiol.*, **77**, 1046–1050.

Tolhurst, D.J., Movshon, J.A. & Thompson, I.D. (1981) The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Exp. Brain. Res.*, **41**, 414–419.

Treisman, A.M. & Gelade, G. (1980) A feature-integration theory of attention. *Cognit. Psychol.*, **12**, 97–136.

Yarbus, A.L. (1967) *Eye Movements and Vision* [translated by Haigh, B.]. Plenum Press, New York.