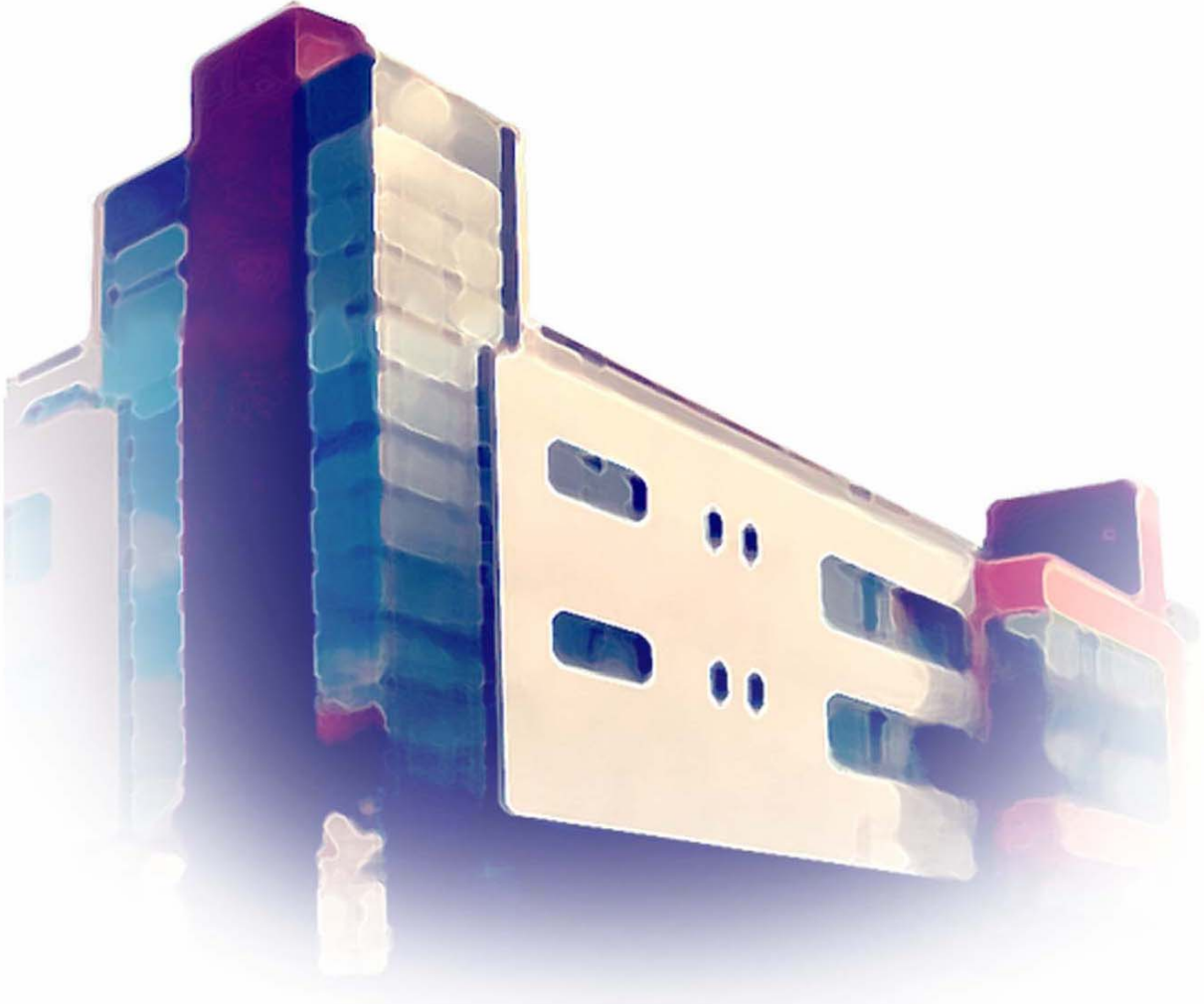


Alper Aık

*Effects of Contrast and its Modifications on
Fixation Behavior during Free Viewing of Images
from Different Categories*



PICS

Publications of the Institute of Cognitive Science

Volume 2-2010

ISSN: 1610-5389

Series title: PICS
Publications of the Institute of Cognitive Science

Volume: 2-2010

Place of publication: Osnabrück, Germany

Date: September 2010

Editors: Kai-Uwe Kühnberger
Peter König
Sven Walter

Cover design: Thorsten Hinrichs

Effects of contrast and its modifications on fixation behavior during free viewing of images from different categories

Alper Aık

Master Thesis submitted for the purpose of the completion of Cognitive Science
Master Program at the University of Osnabrück

2006

First Supervisor: Prof. Dr. Peter Konig
Second Supervisor: Selim Onat, M.Sc.

Table of Contents

1. Outline

2. Introduction

- 2.1. Overt attention and Top-Down Vs Bottom-Up mechanisms
- 2.2. Differences between image categories
- 2.3. Luminance Contrast: Causal or correlational?
- 2.4. Critiques
- 2.5. Motivation of the present study

3. Methods

- 3.1. Subjects
- 3.2. Stimuli
- 3.3. Eye tracking apparatus
- 3.4. Experimental paradigm
- 3.5. Data analysis

4. Results

- 4.1. Basic image statistics
- 4.2. Basic fixation statistics
- 4.3. Effect of luminance contrast
- 4.4. Time course of LC effect
- 4.5. Effect of texture contrast
- 4.6. Effect of modifications
- 4.7. TC change at fixations
- 4.8. TC change - modification effect relationship

5. Discussion

- 5.1. Categories, spatial frequency and LC
- 5.2. When does time matter?
- 5.3. Positive LC modifications
- 5.4. Negative LC modifications
- 5.5. When does TC guide eye movements?
- 5.6. What guides the deployment of attention during viewing of natural images?
- 5.7. The neural basis of Saliency Map

6. Conclusion

7. References

8. Acknowledgements

9. Supplementary Figures

1. Outline:

During viewing of complex natural scenes, eye movements bring portions of the scene into the fovea, where the spatial acuity is highest. This selection is not random, and known to be related to some statistical properties of images. Reliably, luminance contrast has been found to be higher at locations selected as targets of fixations. The contribution of luminance contrast to overt attention changes according to the image category of the scene. To tackle the issue of contrast-attention relationship, we recorded eye movements while human subjects were free viewing unmodified or contrast-modified complex scenes from different categories: natural environments, urban scenes, fractals, and faces. In a recent report Einhäuser and König (2003) argued against a possible causal role of contrast for the selection process, showing that both major decreases and increases in local contrast of natural scenes draw the attention to these regions. The present study first improved and then extended this paradigm to the three other categories mentioned above. Additionally, another group (Parkhurst and Niebur, 2004) argued that both increases and decreases in luminance contrast could lead to unintended increases in texture contrast, a second-order contrast. This can potentially explain the effects of modifications on fixation behavior. Here we checked if this holds for other image categories too. We report that a) luminance contrast is higher at fixated points for each category, but the size of effect changes with category and spatial frequency, b) high positive modifications attract attention in the case of natural scenes, urban scenes, and faces, but not in the case of fractals, c) slight negative modifications have repelling, and high negative modifications have attracting effects for natural scenes, but have no significant effect for other categories, d) unintended texture increase at fixations explains the effect of modifications to a large degree in the case of natural and urban scenes, and to a lesser extent in the case of faces, and fails to do so in the case of fractals. We discuss the results as showing the simultaneous operation of bottom-up and top-down mechanisms in the brain. We argue that the former explains the contribution of luminance and texture contrasts to a saliency map, whereas the latter is related to effects dependent on scene content.

2. Introduction:

2.1 *Overt attention and Top-Down Vs Bottom-Up mechanisms*

Human eyes make around three saccades per second during daily life to bring portions of a visual scene to the fovea. The part of the visual field falling onto the fovea receives disproportionately more cortical processing compared to the periphery and hence is represented with highest resolution (Motter and Belky, 1998). The selection of fixation targets during the analysis of complex natural scenes is not a stochastic process but is guided by overt visual attention to sample the relevant information from the environment. But the exact principles by which fixation selection takes place are still not clear.

There are two classes of candidate mechanisms for guidance of overt attention. On the one hand there are top-down control mechanisms, which are goal driven and task specific. In an early and classical study (Yarbus, 1967), in different trials subjects were shown the same picture, but were told to accomplish a different task (e.g. memorizing the objects in the scene or thinking about a scenario related to the scene). The differences between eye movements as people scan the image reveal how the crucial parts of the image change according to the aim of the task. This line of research has so far made a long way to uncover the relationships between eye movements and simultaneous goal-driven cognitive processes (Hayhoe and Ballard 2005). On the other hand, there are bottom-up, stimulus-driven mechanisms, which are based on the properties inherent in the natural scene itself. In this case statistics of the scene guide the eye movements. Locations, which have certain statistical peculiarities, are thought to attract or repel overt visual attention. Although it is common sense to explain the observed patterns of fixations on natural scenes as interplay of these two mechanisms (see e.g. Navalpakkam and Itti, 2005), currently there is not much consensus about the relative contributions of the two (Tatler, Baddeley and Gilchrist, 2005; Einhäuser and König, 2003; Parkhurst and Niebur, 2004).

Recent modeling studies (Itti and Koch, 2000; Parkhurst, Law and Niebur, 2002) favored a bottom-up approach. This line of research generally postulates the idea of a “saliency map” (Koch and Ullmann, 1985). These maps display the salient, attention attracting portions of a scene in a retinotopical manner. For the extraction of such a map from a visual scene, first the

individual properties like luminance or color contrast are computed separately and then combined for the final feature-independent map (Itti and Koch, 2000; Peters, Iyer, Itti and Koch, 2005). Accordingly, the salient points on this map will draw overt attention. Eyes will be guided to most salient points sequentially, following the saliency order. The finding that certain feature contrasts (Parkhurst et al., 2002; Krieger et al., 2000), especially luminance contrast (Reinagel and Zador, 1999), are found to be higher at fixation locations compared to the other parts of the scene, is generally used as evidence for saliency map accounts.

2.2 Differences between image categories

‘Natural image’ is a term used for a huge variety of scenes, including urban scenes, faces, or scenes from natural environments, without any man-made objects. Are the statistics important for visual processing the same for pictures coming from such different image categories? Torralba and Oliva (2003) analyzed the spectral signatures of scene types, as specific and diverse as mountains, fields, beaches, streets and high buildings, and concluded that these second order statistics can be used for image classification. If the images were pooled into superordinate classes as man made and natural, especially the power in horizontal and vertical orientations led to a clear separation of the two classes. Using a bottom-up approach Parkhurst and colleagues (2002) tracked the eye movements of human subjects as they viewed images from four categories, namely home interiors, natural landscapes, city scenes, and fractals. The investigated features contributing to their saliency map were orientation, color and luminance contrasts. The authors first generated combined saliency maps as a linear sum of single feature saliency maps, and compared fixated points to randomly chosen points in terms of saliency on the combined map. The saliencies at fixated points were significantly higher than those at random points. Then for each database, the fixated point saliencies on combined map were compared to saliencies at single feature maps. The contributions from separate channels varied between categories. Differences between image categories might have significant implications for both top-down and bottom-up accounts of visual attention, given that the gist of a scene can be registered before the local details (Davon, 1977). In the case of bottom-up accounts, varying statistics might change the weights of individual feature channels during the extraction of the final saliency map. On the

other hand, after the scene content is registered top-down mechanisms could guide eye movements effectively to objects (e.g. if a library, look at books) and locations (e.g. if an office, scan the surface of tables) informative for such images. Hence a comprehensive account of overt visual attention must take the differences between image categories into account.

2.3 Luminance Contrast: Causal or correlational?

Does the fact that various types of contrast are higher at fixation points necessarily mean that it is these features that causally attract attention? Goal-driven processes such as visual search can also lead to the selection of locations possible to describe with low-level features (Tatler et al., 2005). Our group recently tackled this issue by the introduction of a novel eye tracking based paradigm (Einhäuser and König, 2003). We chose luminance contrast (LC), since it was the most thoroughly studied contrast, and because it was always found to be higher at fixation points than at other parts of the image (Itti and Koch, 2000; Parkhurst et al., 2002). Unmodified or locally LC-altered natural images with an absence of any man-made objects were shown to a group of subjects. The subjects were only told to ‘study the images carefully’, to avoid any task-specific bias in the eye movements. In the case of altered images, at randomly chosen points LC was increased (positive modification) or decreased (negative modification) without introducing any artificial boundaries. The classical saliency map model predicts an increase in fixation to the LC increased regions and a decrease in the case of negative modifications. This prediction was only partly confirmed. The positively modified regions received more fixations than expected by chance, but the negative modifications acted likely attractive for the visual system as well. Recently the results are replicated and generalized to monkeys as well (Einhäuser et al., 2006). This is contradictory to the saliency map theories and shows that although luminance contrast might be correlated with fixations, the relation is nonetheless not causal.

2.4 Critiques

Parkhurst and Niebur (2004) confirmed the findings of the above-mentioned study (Einhäuser and König, 2003), showing that their computational model, which chooses locations exclusively

according to luminance contrast failed to replicate the results from the modified images. They nevertheless criticized the methodological aspects of the study, some of which are corrected in the current report. In the original study (Einhäuser and König, 2003) the same scene was presented 30 times, each time for 8 seconds. Since only 8 images were used, this could bias the participants to try to memorize each detail in the scene, becoming especially tuned to modifications. Here we report similar results with less and shorter presentations, and a larger image database. Another major concern was the possibility of unintended contrast increase in the case of negative modifications. To change the contrast locally, the difference between the intensity of a pixel and the global intensity of the image was weighted according to the modification level under consideration and added to or subtracted from the value of the pixel. Parkhurst and Niebur (2004) argue correctly that since the luminance in the image is not uniformly distributed (especially in natural images where horizon and sky can be seen), global mean is not a good estimate for the local mean. If negative modifications fall on high or low luminance locations, introducing patches of average luminance would increase the contrast in the region, rather than decreasing it. Even though these are minor details, improvement of the modification technique is necessary to conclusively argue that luminance contrast is not causally related to attention.

In their same study Parkhurst and Niebur (2004) devised another model to better account for the results adding a second order saliency map to represent texture contrast (TC) as well. TC refers to variations of LC, or of other texture signaling elements (Johnson, and Baker, 2004). Unlike first-order contrast (i.e. luminance contrast) judgments, where two points must be observed, for TC the observation of at least three points of the image is necessary (Schofield, 2000). Parkhurst and Niebur (2004) showed that, since both increases and decreases in LC modifications lead to increases in texture contrast increases, the results of Einhäuser and König (2003) can still be explained purely by bottom-up mechanisms. It is worth mentioning that in their model, the contribution of TC was ten times higher than the contribution of LC, further questioning the role of LC in deployment of attention.

2.5 Motivation for the present study

This study aimed at first improving the methodology used by Einhäuser and König (2003) in the light of the suggestions by Parkhurst and Niebur (2004) and then at extending it by using four different categories of images. Together with natural environment pictures similar to those used in the original study, urban scenes, faces and fractals were used as well. Additionally a detailed analysis of modification effect – unintended texture increase is provided.

3. Methods:

3.1 *Subjects:*

The subjects were 14 (5 female) university students, ages 21-27. They were granted with either 20 euros (\approx 20 dollars) or extra course credits for their participation. All had normal or corrected-to-normal vision, and were naïve to the purpose of the study. Written consent was obtained from all of them. The experimental procedure conformed to the national guidelines for human experiments.

3.2 *Stimuli:*

The basis images subjected to contrast modifications (see below) were chosen from four categories (for representative examples see Fig. 1 a-d). The images in the first category (*'naturals'*) depicted natural scenes like forests, open landscapes, with an absence of any man made object. These were taken for the purpose of the original study (Einhäuser and König, 2003), of which this one is a continuation. These images subjectively resembled scenes recorded with a camera mounted on the head of a cat (Kayser, Einhäuser and König, 2003). Images from the second category (*'manmades'*), on the other hand, depicted either urban scenes like streets or house exteriors, or man made objects like barrels or hats. These were part of the 'McGill Colour Calibrated Image Database' (Olmos & Kingdom, 2004). The third category (*'fractals'*) consisted of software generated fractal pictures taken from a World Wide Web fractal database (<http://www.cnspace.net/html/fractals.html>, chaotic n -space network, 1999). The last category (*'faces'*) included frontal, close up shots of faces, taken especially for the purpose of this study. They were taken with a high-resolution camera (Sony DSC-V1 Cyber-shot, Tokyo, Japan) in enclosed public places (e.g. university library) using the available light in the room. All images from all categories were downsampled or cut to a resolution of 1024*768 and converted to grayscale using the built-in MatLab (Mathworks, Natick, MA, USA) function *rgb2gray*. For face pictures care was taken to fill the monitor with the face in the vertical direction as much as possible. In the end 16 images were selected from each of the four categories, making up a total of 64 basis images.

To introduce local contrast changes, we followed the original study of Einhäuser and König (2003) with some improvements. Briefly, random points (x_i, y_i) were chosen and two-dimensional Gaussians with

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

were centered on each of them (λ denotes the standard deviation). The maximum over these was taken to have the mask $G(x, y)$, which is used to modify the pixel intensity at each image point $I_0(x, y)$. For this, Einhäuser and König (2003) used the following equation for any modification level α :

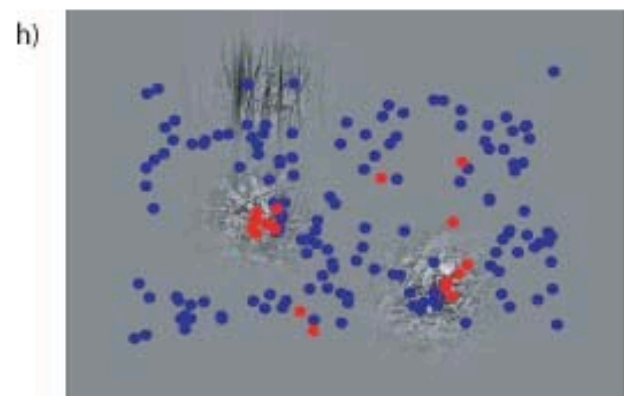
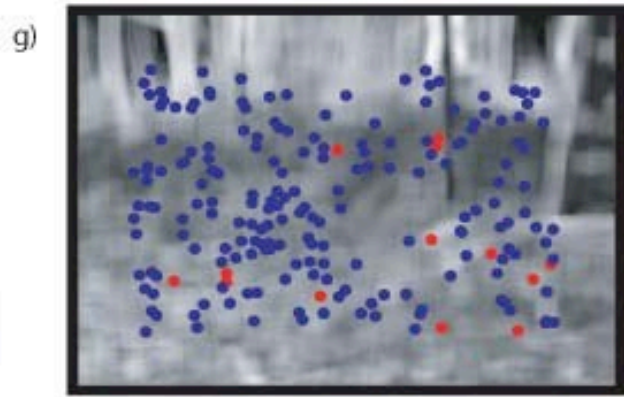
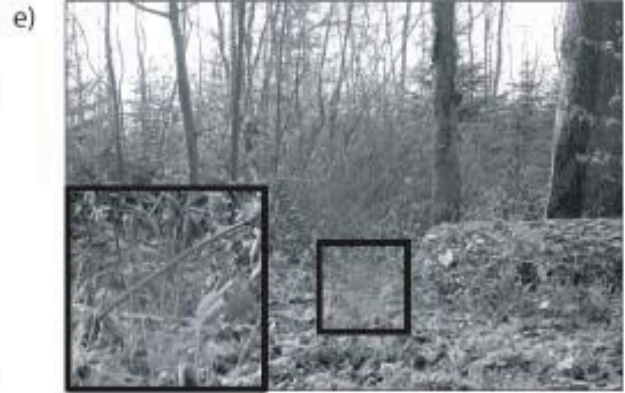
$$I(x, y) = I_0(x, y) + \alpha G(x, y) * [I_0(x, y) - \langle I_0 \rangle]$$

where $\langle I_0 \rangle$ denotes the mean intensity of the whole image. This means that a portion of the difference between the mean intensity of the image and the intensity of a pixel was added to or subtracted from the intensity of this pixel. Parkhurst and Niebur (2004) noted that global mean intensity of the image is not necessarily a good estimate for the local mean, which would be important for local contrast. They argued that in the case of negative manipulations, regions of average luminance would be introduced into higher and lower luminance regions, increasing the contrast in these regions, instead of decreasing it. This problem can be solved using the local mean instead of the global mean. In the present study we first computed the local mean map of the image:

$$\hat{I} = I \otimes G_{unit}$$

Here, \hat{I} is the result of convolution of image I with a Gaussian of unit integral and of the same size as used in the modification. It gives a map of the local mean values in a matrix of the same size as that of the image. Accordingly, the modification equation becomes:

$$I(x, y) = I_0(x, y) + \alpha G(x, y) * [I_0(x, y) - \hat{I}(x, y)]$$



Thus we altered local contrast faithfully without introducing changes in the local luminance. We have randomly chosen three points to center the Gaussians with $\lambda = 80$. These points were at least 160 (2λ) pixels away from each other, 160 pixels (2λ) from the vertical image boundary, and 334 ($1/4$ of the horizontal image size plus 1λ) pixels from the horizontal image boundary. The higher limitation in the case of horizontal direction made sure that modifications did not fall into the uniform background regions in the case of faces. We used modification levels from -0.6 to $+1.0$ (see Fig. E and F for the two maximum modifications, namely -0.6 and $+1.0$). Together with the non-modified versions, this made a total of 9 modification levels for each basis image.

The images were displayed at 1024×768 pixel resolution and 75 Hz refresh rate on a 21" CRT monitor (SyncMaster 1100DF, Samsung Electronics, South Korea), which was located 80 cm away from the subject. The stimuli thus covered $28 \times 21^\circ$ of the visual field.

3.3 Eye tracking apparatus:

Eyelink II (SR Research, Ontario, Canada) was used to record eye movements. It uses infrared pupil tracking and corneal reflection based tracking at 250 Hz with a spatial resolution of 0.01° , and is able to compensate for head movements. In this study, monocular eye position data were sampled with infrared-based tracking only.

To calibrate eye position and to validate this calibration, subjects were asked to make saccades to the nine fixation spots on the screen. This procedure was repeated at the start of tracking a couple of times to warm up the subjects. If the mean calibration error for any eye was below 0.24° in two successive validations, tracking of that eye started. Additionally, after each single stimulus

Fig. 1. Stimuli and data analysis methods. a) to d): From up to down, representative examples from 'naturals', 'manmades', 'fractals', and 'faces' categories, respectively. e) Maximum negative modification level (-0.6) example for the basis image in a). The region in the small square shows one of the three modifications in the image. The larger square is the zoomed version of the same region. The length of the patch corresponds to two standard deviations of the Gaussian. Neither the small square nor the zoomed region in the larger square was shown to subjects. f) Same as in e), but for maximum positive modification level ($+1.0$). g) The contrast map of the basis image in a). The black borders (20 pixels from each side) show the regions where contrast cannot be calculated faithfully, because of the size of the patch (40×40 pixels) used for contrast computation. Overlaid are the actual (red) and control fixations (blue) of a single subject. Actual fixations come from the presentation of the basis image, and control fixations from the presentations of the rest of the basis images in the same category. h) The pixelwise difference between maximum positive modification and the basis image, highlighting the three modified regions. Uniform gray areas are zero regions. Overlaid are the actual (red) and control (blue) fixations of a single subject. Actual fixations come from the presentation of the maximum positive modification image, and the control fixations from the presentations of the other modification levels of the same basis image. Please note that g) and h) are not from the stimuli shown to subjects, but are for descriptive purposes only.

presentation a fixation spot appeared in the middle of the screen. The error of the eye movement to this spot was used to correct for the drift of the tracker. If this error was larger than 1° , calibration and validation was repeated till the error criterion (i.e. $< 0.24^\circ$) was met again.

Fixation locations and times were calculated online by the eye tracker. The parameters to define the saccades were $8000^\circ/\text{sec}^2$ for acceleration threshold, $30^\circ/\text{sec}$ for velocity threshold, 0.1° for deflection threshold. This state had to be maintained for at least 4 ms. The extraction of saccades left the fixations over, which are used for further analysis.

3.4 Experimental Paradigm:

During the course of the experiment each subject saw a total of 576 pictures (16 pictures per category \times 4 categories \times 9 modification levels) each of which is displayed for 5 seconds on the screen. The presentation order was completely pseudo-randomized, but two modifications of the same basis image were never shown consecutively.

Given that eye tracker starts to give discomfort after about 45 minutes, it was not possible to show whole stimuli set in one session. We divided our stimuli set into 4 blocks of 144 pictures. Each of these blocks was shown in one session, conducted on different days. For none of the subjects the course of the whole experiment exceeded one week.

At the start of each session the subject was taken to the darkened eye-tracking room. The only instruction was to ‘study the images carefully’, avoiding any task-specific bias. After the eye-tracker was installed on the head, the subject was brought to the front of the display monitor. The middle of the screen was at about the eye level of all subjects. With the approval of the subject calibration and validation started as explained above. In each session after 72 image presentations a five-minute break was given, and the eye tracker was removed from the head. This ensured that the eye tracker was never on the subject’s head for more than 30 minutes, how long the calibration might take. After the break tracking was initiated with calibration and validation again. Single blocks did not exceed a total of 45 minutes. At the end of the fourth session the subject was informed about the purpose and details of the study.

3.5 Data Analysis:

The first fixation from each stimulus presentation was discarded, since these points were always the artifacts of the fixation spots between presentations. For all the following analyses, only those fixation points were taken into account that were 80 pixels away from the sides of the screen, to get rid of border effects.

We first wanted to know if in unmodified images luminance contrast (LC) was correlated to fixations. For LC, we employed the definition of Reinagel and Zador (1999), according to which LC at a given point is the standard deviation of luminance in a square region of the image around this point, divided by the mean intensity of the image. In the present study the length of the square region was 40 pixels, or $1,1^\circ$. Parkhurst and Niebur (2003) found the most striking effects of LC for $1^\circ \times 1^\circ$ patches, which is very close to our choice. Since standard deviation is defined

as $\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$, where brackets denote the mean, the luminance contrast map can be computed with the following formula in a computationally effective manner:

$$(\sqrt{(I^2 \otimes p) - (I \otimes p)^2}) / \langle I \rangle$$

Here I denotes the image, \otimes the convolution operation, p a 40 by 40 pixels patch with unit integral, and $\langle I \rangle$ the mean of the image (see Fig. 1 G for an example).

For each image the fixations taken from the presentation of this image (actual fixations) were compared to the fixations gathered from the rest of the images in the same category (control fixations). This controls for any viewing biases and is extremely important, given that we have a *faces* category. Taking random points, instead of fixation from other images, we would end up in uniform, zero-contrast regions as control points, which would confound our results. This approach is also supported by the non-uniformity of the average LC maps (Tatler et al., 2005). Furthermore, in the cases where only those fixations were taken, which fell into a certain time window of the presentation, the same temporal criterion was used for the control fixations as well. The same kind of analyses was performed on the low-passed versions of the images as well. Einhäuser and König (2003) failed to find a LC-fixation relationship when the contrast was

computed on original images, but this relation became obvious when low-passed images were taken. We compared the contrast at actual and control fixations with two-sided sign tests.

The next concern was if the modifications had an effect on overt attention. The direction of such an effect is very important for testing models of saliency, in the sense that it can show if luminance contrast has a causal role in attracting or repelling attention. Such a causal role would be supported in a scenario, where positive modifications receive more fixations, and negative modifications less fixations than an appropriate control condition. To compare the distribution of modifications on the actual fixations to a given modification level of a basis image, we took the fixations from the presentations of the same basis image in different modification levels and concatenated them. The distribution of modifications on this trace established the control distribution. The similarity of these two distributions was quantified using a two-sided Kolmogorov-Smirnov (KS) test. From the actual and control modification distributions cumulative distributive functions (CDFs) were computed. KS-test measures the maximum distance between the two CDFs coming from independent samples. But two subsequent fixation points are not independent, because the location of a fixation is effected by the location of the previous one. To solve this problem we computed an effective sample size looking at the autocorrelation function of the locations of the successive fixations in vertical and horizontal directions. Accordingly we decreased our sample size and used it in the KS-tests. For each category and modification level separately, the data from each image and subject was pooled together to check if the modifications had an effect on fixations. Furthermore another measure to quantify the effects of modifications was introduced (see Results), to compare to the obtained KS-results (Einhäuser et al., 2006).

The following step was to check if there was a relationship between the texture contrast (TC) of the image and LC or LC modifications. TC is the local variance in LC or other texture defining elements. It is generally studied together with LC as a second-order vision mechanism, compared to first-order luminance based mechanisms (Schofield, 2000). TC was enough to explain the results of the original study (Einhäuser and König, 2003), as shown by Parkhurst and Niebur (2004). Modifying the LC in either direction was increasing the TC, and this could potentially explain why modifications of either type were attractive for visual system. To address this issue, the change in TC, which is an unintended result of LC alterations, was computed. We defined TC as the standard deviation of LC in a square region, divided by the mean LC of the image, to

which the region belongs. One side of the square was 200 pixels unless otherwise stated. Taking into account the size of LC calculation patch, for the computation of TC at any given point, luminance values from a 240 x 240 patch are considered. The texture contrast map can be calculated in analogy to the luminance contrast map calculation above, replacing the image with the luminance contrast map.

4 .Results

4.1 Basic Image Statistics

There were differences between categories in terms of their first and second order statistics. The average spectral signatures are category specific (Fig. 2), and replicate the previous findings (Torralba and Oliva, 2003). As a result of zero-padding the power spectra had square shapes, unlike the rectangular images. Relatively high power in vertical and horizontal orientations in the case of *manmades* is obvious. The averages of local feature values for each image were also

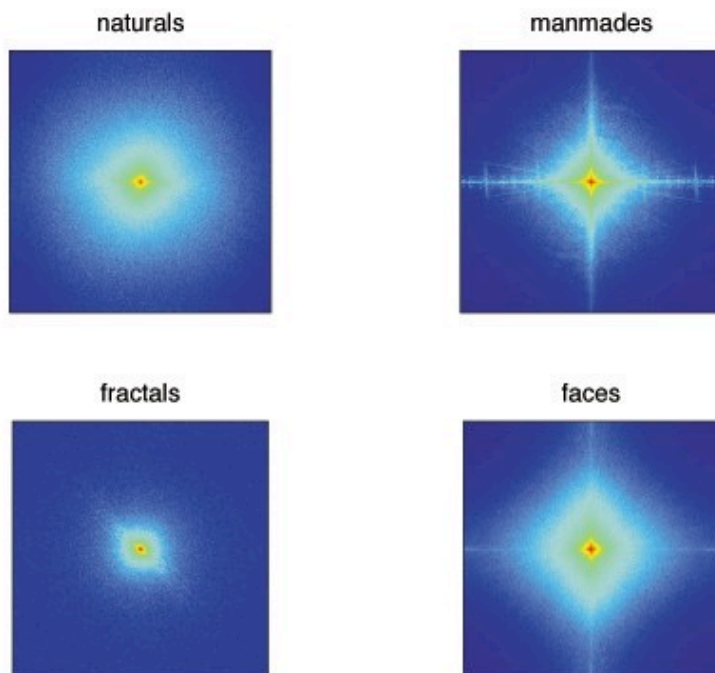


Fig. 2. The average two-dimensional spectral signatures of the four categories used in the study. The logarithmic color code indicates higher power from blue to red. The spectra are square-shaped, unlike the rectangular images, due to the zero-padding.

calculated. This was done through taking the mean over the control fixations of that image. This ensured that we controlled for general viewing biases, like the central bias. This approach is especially preferable in the case of faces, since there we had large uniform background regions, which were never fixated. The means and standard deviations of the three features under consideration are given in Table 1. Due to these differences, an absolute difference between actual and control contrasts of one category could not be compared to that of another one. This

even holds for two images in the same category, especially in *fractals*. Therefore to account for the differences between categories, we applied a difference normalization procedure as explained below.

Table 1. Means and standard deviations (in brackets) of luminance (scaled to be between 0 and 1), luminance contrast (LC) and texture contrast (TC)

Category	Luminance	LC	TC
Naturals	0.47 (0.04)	0.29 (0.07)	0.06 (0.02)
Manmades	0.46 (0.10)	0.22 (0.08)	0.12 (0.04)
Fractals	0.51 (0.14)	0.34 (0.18)	0.13 (0.08)
Faces	0.53 (0.08)	0.15 (0.04)	0.07 (0.02)

4.2 Basic Fixation Statistics

The fixation durations and number of fixations were similar to those found in the literature. The average fixation durations were 283, 277, 295 and 290 milliseconds for *naturals*, *manmades*, *fractals*, and *faces*, respectively. A two-way 4 (category) x 9 (modification) ANOVA failed to find any difference between modification levels ($p > 0.05$). But the effect of category was significant ($p < 0.05$). Post-hoc analysis with t-tests revealed that this was due to the difference between *fractals* and *manmades*. Neither the interaction nor the other pairwise comparisons were significant. The tendency to keep a fixation longer in the case of *fractals* replicates the findings of Parkhurst and Niebur (2003).

Figure 3 shows the smoothed probability distributions for fixation location for each category. For the calculation of these maps, the fixations to each image in a single category were put on top of each other. This map was convolved with a Gaussian (standard deviation = 30 pixels). Again we confirm the findings of Parkhurst and Niebur (2003), showing that the central tendency is more

profound during viewing of *fractals*, compared to *naturals* and *manmades*. In the case of *faces*, the typical triangular scan pattern (two eyes and the mouth) is obvious.

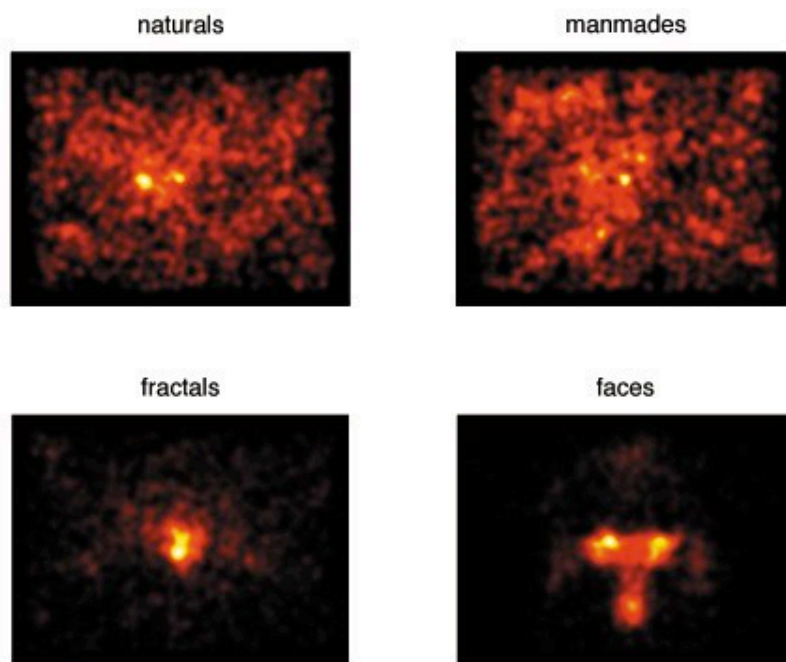


Fig. 3. The probability maps for each category. From black to yellow the probability for fixation increases.

4.3 Effect of Luminance Contrast

The first fixation-based analysis concerned the comparison of actual and control contrasts in unmodified images. The actual contrasts to a given image are the luminance contrast (LC) values at the fixations to this image (red dots in Fig. 1g). The LC values at the fixations to the other images in the same image category (blue dots in Fig. 1g), but again laid on the same image established the control contrasts. The difference between the two was tested for significance with a sign test. Computing the LC at the stimulus presentation resolution we got highly significant effects of LC for *fractals*, *manmades*, and *faces* (all p-values < 0.001), but not for *naturals*. Einhäuser and König also failed to find an effect of LC with their natural images, when it was calculated at full picture resolution, but they were able to do so, when they used low-pass filtered images. We repeated our analysis on low-pass filtered images, using different cut-off frequencies.

The chosen cut-off frequencies were 1.38, 0.69, and 0.38 $\text{cyc}/^\circ$. If LC was defined on the low-pass images with any of these cut-off frequencies, we get highly significant results for

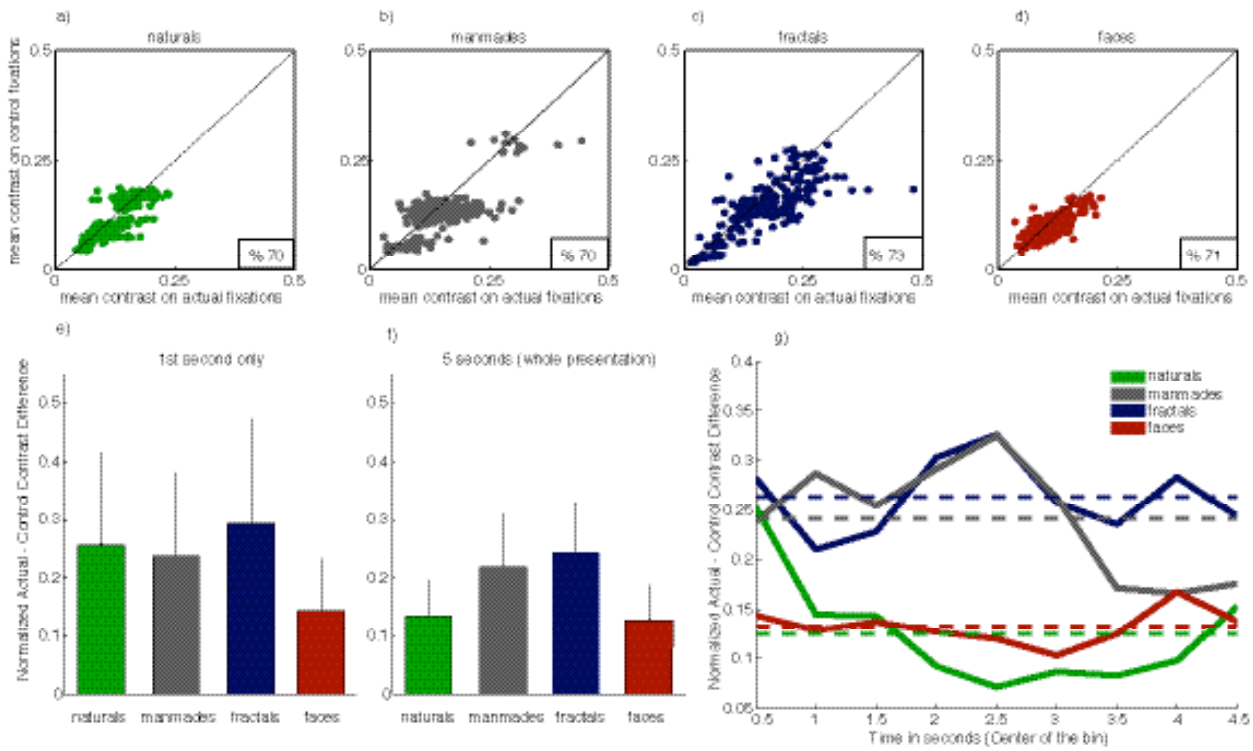


Fig. 4. Effects of luminance contrast on fixation behavior. In all the subfigures contrast is computed on low pass images with cut-off frequency 0.69 cycles per degree. For contrast computations on original images and on low pass images with different cut off frequencies, please refer to the text. a) to d): Mean contrast on actual versus control fixations for each image and subject. Points below diagonal imply higher mean contrast on actual fixations. The rectangular insets give the percentage of points falling below the diagonal. The differences in mean contrast of the images from different categories (see text) are obvious. e) Normalized mean contrast difference between actual and control fixations in the first second of the presentation. The difference between mean contrast on actual fixations and control fixations was divided by the mean contrast on control fixations, for each image and subject separately. For both actual and control conditions only those fixations were chosen which happened in the first second of the stimulus presentation. The bars are the grand means of this analysis for all subjects and images. Error bars are for the standard error of measurement for different pictures, averaged over subjects. Actual versus control comparisons with sign test revealed highly significant (all p -values < 0.001) results for all categories. f) Same as in e), but for the whole presentation (5 seconds). Sign test comparisons are still highly significant for all categories (all p -values < 0.001). g) Shifting window analysis for normalized actual minus control difference. A one-second long window was shifted with a window overlap of 0.5 seconds and the same analysis as in preceding subfigures was performed. The solid lines are the resulting curves. The mean over time is plotted as dashed lines, using the same color code.

all the categories (all p -values < 0.001 , except *faces* at 0.38 $\text{cyc}/^\circ$, where $p < 0.01$). We chose 0.69 $\text{cyc}/^\circ$ for further analysis, since at this cut-off frequency, the difference between actual and control contrasts were very high, and the percentage of subject-picture combinations, where actual contrast was higher than control contrast was very similar for each category (Fig. 4 a-d).

As obvious from Table 1 and Figure 4 a to d, there were major differences in the mean LC of our image categories. This difference made it hard to compare the effects of LC for one category to another. Hence we normalized the difference between mean actual contrast and control contrast, dividing it by control contrast. In other words this normalization divides the actual contrast by control contrast, and then subtracts 1 to give a ratio-based difference. This ratio differences were 0.13, 0.22, 0.24, and 0.13 for *naturals*, *manmades*, *fractals*, and *faces*, respectively (Fig. 4 f).

4.4 Time course of LC effect

About the time course of the LC effect, Reinagel and Zador (1999) report that the first 4 seconds are more crucial than the rest of the stimulus presentation. Following reports come up with different time windows (first 3 seconds in Parkhurst et al., 2002; 2nd to 4th seconds in Einhäuser and König, 2003). But Tatler and colleagues (2005) argue that this effect of time is due to methodological problems. The first fixations to any image might be different in terms of location (central) or the distance between them (closer), giving rise to a temporal bias. In the present study, actual fixations from certain time windows are always compared to control fixations from the same time windows, overcoming any temporal bias. The results indicated that influence of time on the LC effect depended on the image category under consideration (Fig 4 e-g). Comparing the 1st second of presentation to the whole presentation (Fig. 4 e and f) shows that LC effect is strongest for *naturals* very early. There is not a clear effect for the rest of our categories. To better understand the effects of time, we performed a sliding window analysis (Fig. 4 g). The window length was 1 second, and the window overlap was 0.5 seconds. We moved both actual and control fixations in time according to their onset to overcome the potential temporal bias mentioned above. For *naturals*, the effect was very high at the beginning, then dropped below the mean of the whole presentation effect, and then started to increase again towards the end of the presentation. For *faces* and *fractals* there was no clear effect of time, and for *manmades* even though the first 3 seconds seemed more important, the decrease after that was not as dramatic as in the case of the *naturals*. Overall, except for *naturals*, we did not find a systematic effect of time, in line with findings of Tatler and colleagues (2005), who used home interiors and urban scenes as stimuli.

Is there a relationship between the mean TC of the image and the attention effect of LC? Since TC refers to the deviations of LC (at least in our analyses) it could have an influence on how much the subjects will prefer high LC regions over low LC regions. Even though there was no effect within categories, the LC effect and mean TC of category averages displayed a linear relationship (Fig. 5). The categories with low average TC (*naturals* and *faces*) show smaller LC effect sizes compared to the categories with high average TC (*fractals* and *manmades*).

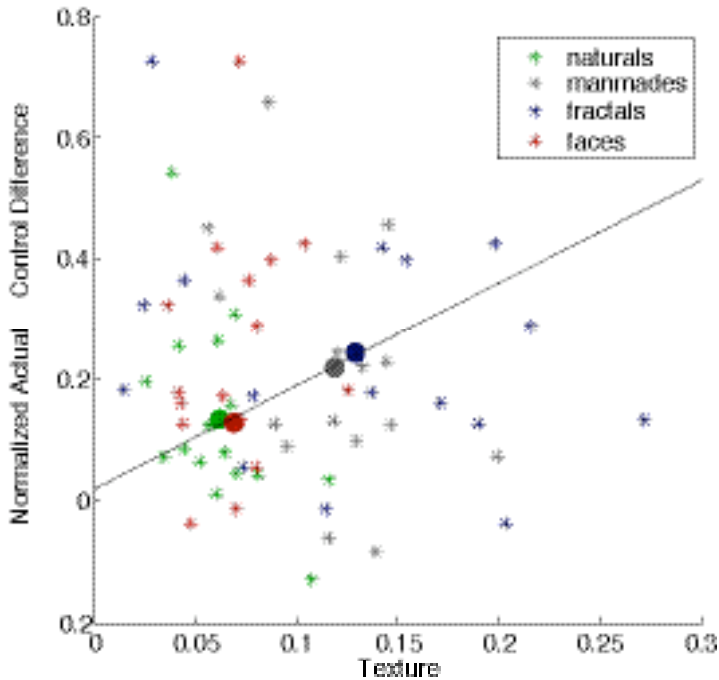


Fig. 5. Effects of mean texture contrast of the images. Texture contrast is defined as the standard deviation of luminance contrast in a 200×200 pixels region. We computed the mean texture of each image on the control fixations for that image. The mean effect of contrast for each image is plotted against the mean texture of images. Filled circles are the category means with the same color code. The black line is the linear fit to these four means ($r^2 = 0.9$, $p < 0.001$).

4.5 Effect of Texture Contrast

Again in the unmodified images, we compared actual TC to control TC, to see if this second-order contrast is also important for fixation selection. The TC difference between actual and control TCs was not significant for *fractals* and *faces* (sign-test, p -values > 0.05), there was a significant but small effect for *naturals* ($p = 0.02$), and a larger effect for *manmades* ($p = 0.009$). The percentage of cases where actual TC was larger than control TC were 58%, 59%, 44%, and 54%, for *naturals*, *manmades*, *fractals*, and *faces*, respectively (Fig. 6 a). We normalized the

actual TC and control TC as in the case of LC effect (see above) (Fig. 6 b). Again we found that the strength of TC effect size was dependent on the image category.

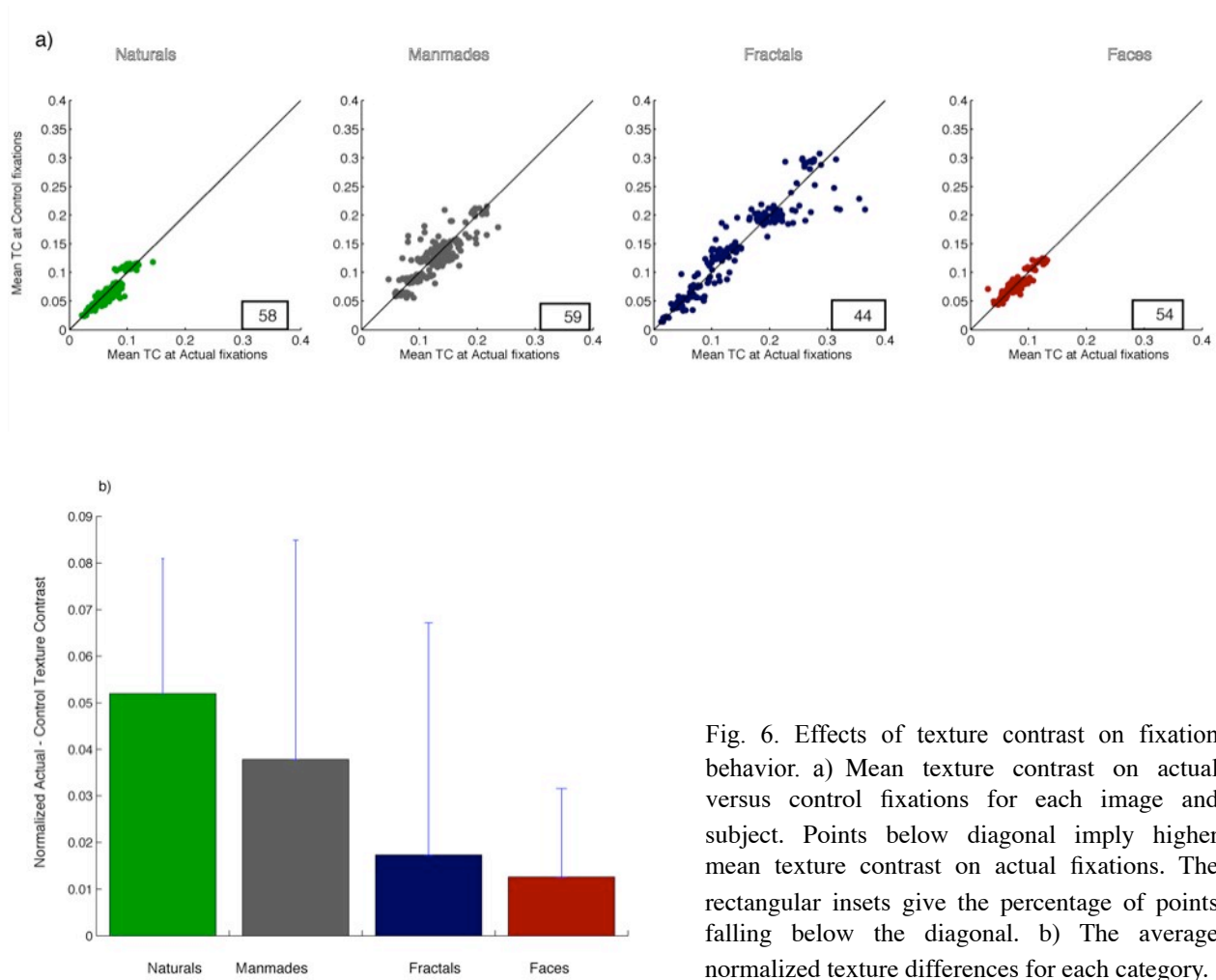


Fig. 6. Effects of texture contrast on fixation behavior. a) Mean texture contrast on actual versus control fixations for each image and subject. Points below diagonal imply higher mean texture contrast on actual fixations. The rectangular insets give the percentage of points falling below the diagonal. b) The average normalized texture differences for each category.

4.6 Effect of modifications

The next step was to quantify to what degree the modifications attracted or repelled the attention, and to check if this was category-specific. In the case of modified pictures, actual fixations to a modified image were compared to the fixations of the same image in the other modification

levels. Since the modified regions are different at each modification this approach provides a very careful comparison, taking into account all possible viewing biases for that single image. We first focused on the maximum positive modification level (+1.0), where we expected to find the most striking effects of our contrast alterations. We looked at each subject and picture combination, and calculated the percentage of cases where the modifications were higher on actual, compared to control fixations (Fig 7 A-D). The percentages were 81%, 66%, 48%, and 75% for *naturals*, *manmades*, *fractals*, and *faces*, respectively. Apparently, for *fractals* our subjects did not look at the areas, where contrast was highly increased, but did so for the rest of the categories, especially for *naturals*.

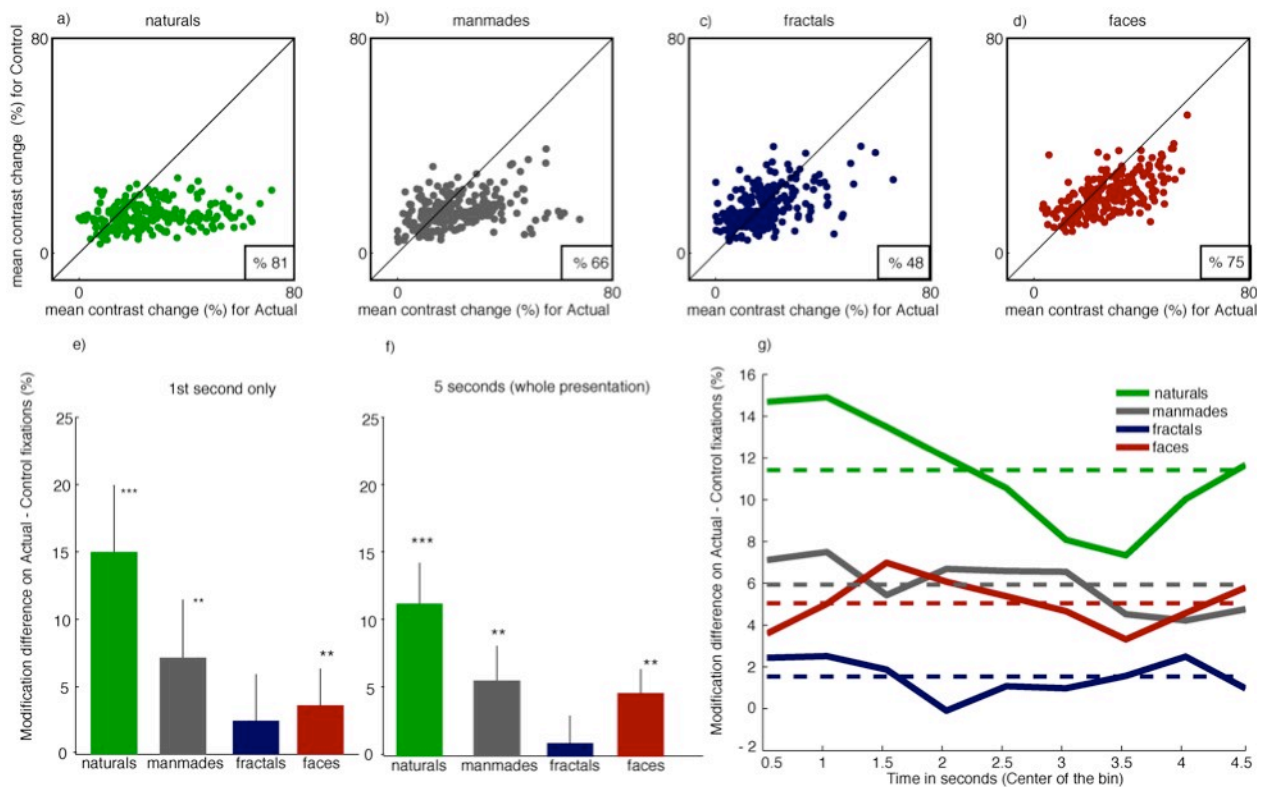


Fig. 7. Effects of maximum positive modification (+1.0) on fixation behavior. a) to d): Mean contrast increase (%) on actual and control fixations are plotted against each other for each image and subject. Points below diagonal imply larger contrast increases on actual fixations. The rectangular insets give the percentage of points falling below the diagonal. e) Mean contrast increase difference between actual and control fixations in the first second. For both actual and control conditions only those fixations were taken into account, which happened in the first second of the presentation. Contrast increase on control fixations was subtracted from the contrast increase on the actual fixations, for each image and subject separately. The bars are for the grand means of this analysis for all subjects and images of the given category. Error bars are for the standard error of measurement for different pictures, averaged over subjects (***, $p < 0.001$; **, $p < 0.01$). f) Same as in e) but for the whole presentation. g) Shifting window analysis for actual minus control contrast change. A one-second long window was shifted with a window overlap of 0.5 seconds and the same analysis as in preceding subfigures was performed. The mean over time is plotted as the dashed lines, using the same color code.

We tracked the effect of maximum positive modification in time, analogous to our LC analysis explained above (Fig 7 E-G). The size of effect stayed constant over the presentation duration for *manmades*, *fractals*, and *faces*, but varied in the case of *naturals*, where we had the largest effect. In this latter case the effect was strongest in the first 1.5 seconds, then started to drop, and stabilized over the mean effect size towards the end of the presentation. The time course of the maximum positive modification effect was also category-specific, and time mattered only in the case of *naturals* just like in the observed LC effect.

We tested the difference between the distributions of modifications on actual fixations and on control fixations with Kolmogorov-Smirnov tests (KS-tests) for each category and modification level. KS-test measures the maximum distance between to cumulative distribution functions (CDFs). We performed this test on our actual and control distributions two-tailed, so that we could find both attracting and repelling effects of modification. We then plotted the maximum CDF differences (Fig. 8). The indifference of fixation behavior to modifications on *fractals* is striking. The *faces* and *manmades* gave similar curves. In the case of negative modifications, there were neither attention attracting, nor repelling effects. However the picture was different for the luminance contrast increased images. Given positive peak modifications equal to or above

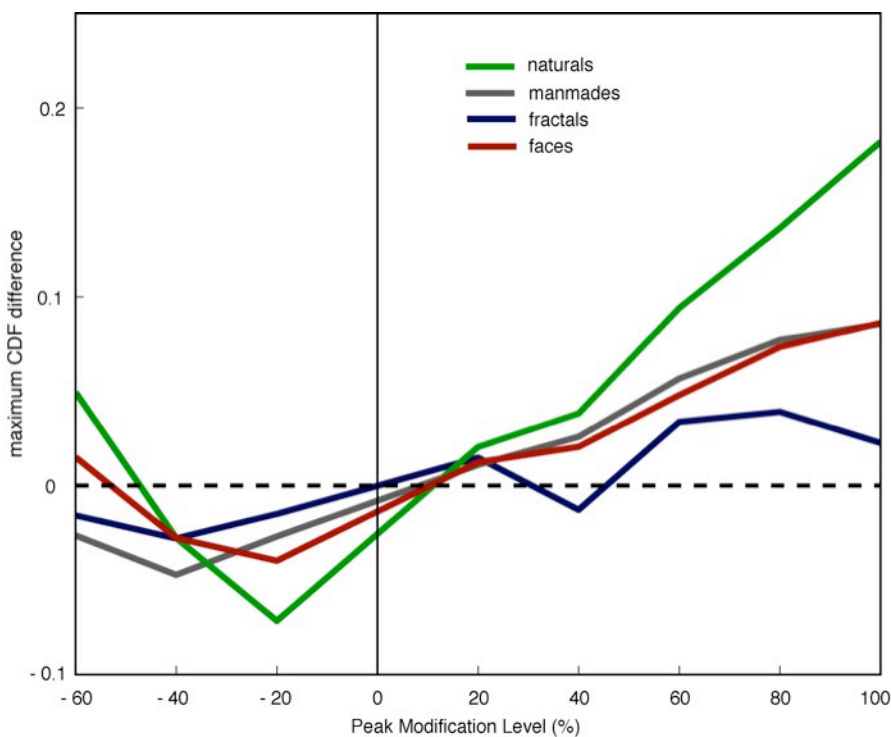


Fig. 8. Maximum CDF differences for all modification levels and categories. All modifications on actual and control fixations were collapsed over images and subjects, but separated according to modification level and category. For each of these actual and control pairs cumulative distribution functions (CDFs) were generated. Here the maximum distance between the actual and control CDFs of each category and modification level is plotted. The plot is scaled such that in cases where modifications attract attention, the difference is positive, and in cases where modifications repel attention the difference is negative. For the significance levels of these differences see text.

60%, the subjects fixated those points more than expected by chance (all p -values < 0.001). The effect increased with the peak modification level. The *naturals* curve behaved similarly on the positive modification side, but the slope of CDF difference change with increasing modification level was larger. The negative side of the curve, on the other hand, did not behave monotonously. In the minimum negative modification level (peak -20%) the modifications repelled the attention significantly ($p < 0.01$), speaking for a causal role of LC in fixation behavior. After the disappearance of this effect at -40% peak modification level, at -60% peak modification level, the effect was reversed, and modifications attracted the attention significantly ($p < 0.05$), replicating Einhäuser and König's (2003) finding.

Following Einhäuser and colleagues (2006) we introduced a saliency measure to assess the effects of the modifications, to check if it would overlap with the KS-test results. For each modified picture the difference between mean actual modification and mean control modification was computed. In the original study (Einhäuser et. al. 2006) median was used instead of mean. But this was not appropriate for us for the following reason. In their study 5 points were chosen to introduce modifications, leaving only small portions of the image unaltered. In our case on the other hand, there were only three modifications, and hence many regions remained intact. Even though we had many fixations to highly modified regions (as obvious from KS-test results) the median modifications we had often were zero, an artifact of large non-modified regions. So we decided to use the mean actual and control modification difference. Like in their study (Einhäuser, 2006) we divided this difference by the modification level of the picture (changing from -0.6 to $+1.0$). Dividing by the modification level ensures that the effect size is not dependent on the peak modification level, enabling us to compare the saliencies from different modification levels. The mean saliencies computed as such, keeping modification and category separate are plotted in Figure 10 A. The resulting curves look nearly identical to the KS-results, legitimizing the further use of this measure in the following analyses (compare to Fig. 8).

Since TC was defined as the standard deviation of LC, it can have a scaling effect for modifications. In an image with low mean TC, the LC modifications can lead to new LC values, which leave the normal range of LC in the intact image. Therefore a negative correlation between the mean TC of an image and the saliency of modifications is possible. We attempted to uncover

this for maximum positive modification saliencies. Even though we found the highest saliencies and lowest mean TC values for *naturals*, and the lowest saliencies and highest mean TC values for *fractals*, we could not show a systematic relationship that would hold for all categories (Fig. 9).

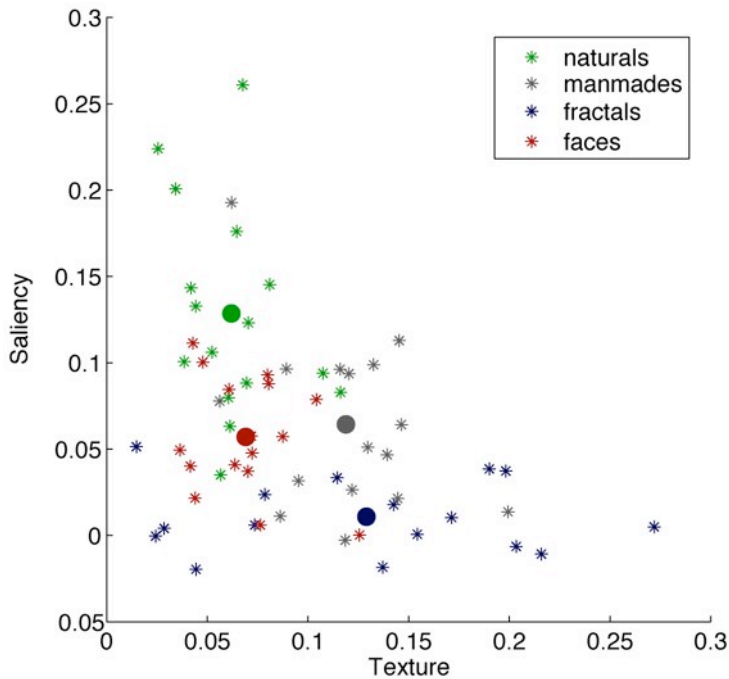


Fig. 9. The mean effect of modifications for each image is plotted against the mean texture contrast. Filled circles are the category means with the same color code. The line fit did not reach significance ($p > 0.05$).

4.7 TC change at fixations to modified images

During viewing of modified images, did the subjects preferentially select those points for fixation, where TC was increased unintentionally as a result of LC modifications? To answer this question, we calculated a texture change index for actual fixations (solid lines in Fig 10 B) to the modified images. For each subject and modified picture we calculated the TC at the actual fixations and subtracted the TC at the same locations in the unmodified version of the same image from that. We further divided this difference by the TC of the unmodified image points, in analogy to the normalizations above. The similarity of TC change at fixations to the saliency of the modifications can be seen by comparing Fig 10 A to Fig 10 B in the case of *natural*, *manmades*, and *faces*. Despite an increase in TC, the *fractals* modifications failed to draw attention.

But given that there is some TC change as a result of LC modifications, the question arises, whether TC change at fixated points is more or less than the overall TC change in the image. If the TC change is the same for both fixated and non-fixated regions we cannot draw conclusions about why the subject chose the modified regions in particular (since the scale of TC is much

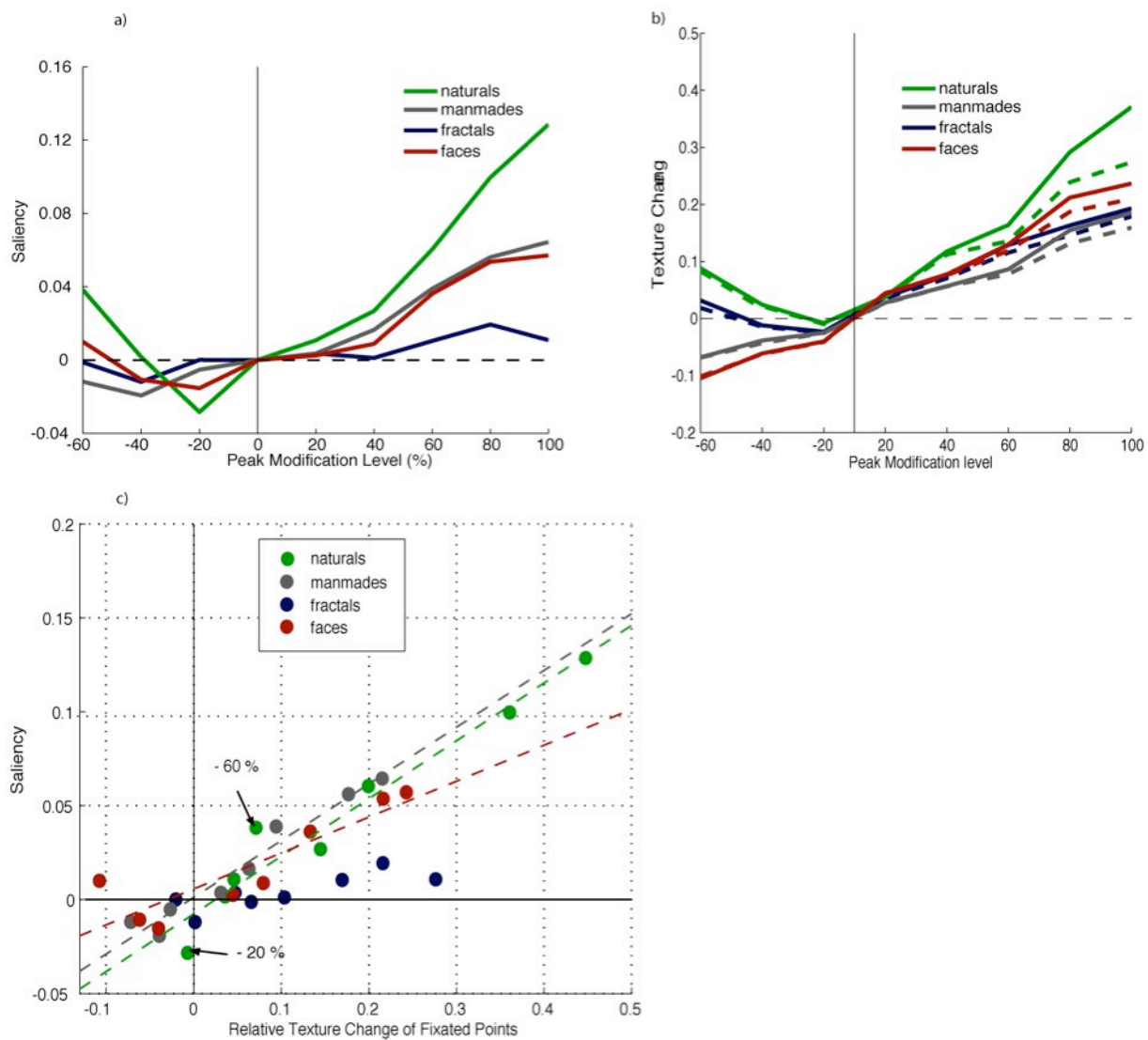


Fig. 10. Effects of unintended texture contrast change on fixation behavior. a) Mean saliency for all categories and modifications. For each modified image, saliency is the difference between the mean modification on actual and control fixations divided by the modification level. Collapsing over subjects and images in the same category, and keeping category and modification level separate, grand means are computed. This measure gives a perfect match to the KS-test results (compare to Fig. 5), and hence used in the following texture analysis. b) The relative change in texture on actual (solid lines) and control (dashed lines) points. The texture change was computed taking the texture contrast on fixation points to the modified image and subtracting from that the texture contrast, calculated on the

same x and y locations but on the basis image. Further dividing this difference by the original texture contrast gives the relative increase on that point, which is an unintended result of the luminance contrast modification. Taking the mean over subjects and pictures in the same category gives an overall texture change on fixations for each modification level and each category. c) Texture change – saliency relationship. Here we plot the ordinates of the preceding two subfigures against each other (from b) only solid lines are taken). This reveals a linear relationship between texture contrast increase and saliency for some categories. The three dashed lines are the linear fits to data, and color code for category is the same as the markers. For ‘*naturals*’, ‘*manmades*’, and ‘*faces*’ the r^2 values to the fits are 0.94 ($p < 0.001$), 0.96 ($p < 0.001$), and 0.78 ($p < 0.01$), respectively. The fit for ‘*fractals*’ did not reach significance ($p > 0.05$). The two interesting modification levels for ‘*naturals*’, -0.6 and -0.2, are highlighted with arrows.

larger than LC, in our case six fold, the TC of points falling outside the modifying Gaussians were altered as well). Therefore we calculated the same TC change index at the control fixations to the same modified image as well, as a measure of unselective TC change (dashed lines in Fig 10 B). For the three categories with significant effect, the positive modification side of the panel displays a difference between selective and unselective TC change, whereas the curves for *fractals* are fairly similar to each other. The negative modification side of the panel, on the other hand, gives us very similar unselective and selective TC changes for each category. This being true for *naturals* questions whether it was the increased TC, which draws attention to the modifications, since we find such a TC change over the control fixations as well.

4.8 TC change - modification effect relationship

We tested the relationship between unintended TC change and the effects of modification applying linear regression analyses for each category separately. We first plotted the mean TC change at actual fixations for each modification level against the mean the saliency of the same modification (Fig. 10 C). Then using linear regression analysis we fitted lines to see if there is a relationship between the two. For *naturals*, *manmades*, and *faces* the r^2 values to the fits are 0.94 ($p < 0.001$), 0.96 ($p < 0.001$), and 0.78 ($p < 0.01$), respectively. This shows that for these categories, unintended texture increases might have guided the eyes to the modified regions, supporting a causal role of texture contrast in our paradigm. The fit for ‘*fractals*’ did not reach significance ($p > 0.05$). One interesting finding is that in the case of *naturals*, the LC decrease repels attention only if the modification does not lead to an unintended TC alteration (i.e. at –

20% peak modification). Overall the unintended TC increase explains the saliencies of LC modifications at least to some degree and in some of the categories.

5. Discussion

5.1 Categories, luminance contrast and spatial frequency

To study the correlational and causal roles of luminance contrast (LC) we studied the eye movements of human subjects as they viewed LC unmodified or modified images from different categories, namely *naturals*, *manmades*, *fractals*, and *faces*. Our findings from unmodified images confirmed earlier findings showing increased LC on fixated points compared to control points (Reinagel and Zador, 1999; Krieger et. al. 2000) at least in a certain spatial frequency range, and the size of effect was category-dependent (Parkhurst et. al., 2002). We confirm Parkhurst and Niebur's (2003) finding by showing that LC effect is higher for *manmades* than *naturals*. Additionally they had shown that saliency was most important for *fractals*, and argued that this is due to decreased intervention from top-down processes, which might operate in the case of meaningful pictures even in the absence of a specific task. Our finding that LC effect was highest for *fractals* is in line with their result. The fact that in their case contribution from LC to saliency in *fractals* was low can be attributed to differential color content of the images used in the two studies. They found that color contrast was most important for their fractals. Since we had grayscale pictures, LC might have become more important. In the case of *faces* it is not surprising that we found an effect of LC, since the eyes and mouth are high LC regions. If the selection of these regions is a result of a top-down process remains to be tested. It is interesting that we found the effects of LC on full resolution pictures for all categories except *naturals*, for which the effect was apparent only after low-pass filtering of the images. In the latter case one could argue that the low frequencies are important for fixation selection, because the resolution drops dramatically as one moves away from the fovea (Motter and Belky, 1998). This property is of course independent from the category to which the image being viewed belongs. Therefore high frequency components of increased LC in the three other image categories might just be correlated to other image properties, which are actually important for fixation selection. One candidate for such an image property is LC on low-passed images, because in this case we find elevated LC values for all of the image categories we tested. Altogether, the data from unmodified regions adds to a large body of observations, where luminance contrast has been found higher at fixations compared to control points, even though the difference is not large.

5.2 When does time matter?

Tatler and colleagues (2005) argued that previous findings that saliency guides eye movements in the early portions of presentation and that the LC effect disappears thereafter (Reinagel and Zador, 1999; Parkhurst et al, 2002) are just methodological artifacts. According to the authors the early fixations differ from late fixations regardless of the image being viewed, and hence one must take this temporal bias into account during selection of appropriate control points. In the present study we compared the image statistics at fixation points in a certain time window to the statistics of the same image at fixation points from other images, but again in the same time window, overcoming image-independent temporal biases. The effect of LC dropped only in the case of *naturals*, and stayed stable or fluctuated without a clear behavior in other categories. This does not contradict the results of Tatler and colleagues (2005), who used mainly urban scenes and home interiors in their study. Our results indicate that given scenes, where one can see clearly discernable objects or object-like structures (*fractals*), the effect of LC stays stable over time, but drops in the case of our natural environment pictures like we used. In the former three categories, this result questions the hypothesis that bottom-up processes lead attention initially, until top-down processes start to display their effects (Parkhurst et al, 2002). In the case of natural environments with an absence of man-made objects, on the other hand, high LC regions seem to be correlated with attention only in the first second.

5.3 Positive LC modifications

To directly test the hypothesis that LC is causally related to overt attention, we increased or decreased LC locally, without introducing artificial boundaries (Einhäuser, and König, 2003). If LC is causally related to fixation relation, increases should draw attention, and decreases should repel it. The former prediction was validated for peak modification levels 60% and more, for all categories except *fractals*. A positive effect could be seen as supporting the causal role of LC in purely bottom-up models of attention. Alternatively, since modifications of this magnitude might leave the range of contrast fluctuations seen in natural images (Einhäuser et al, 2006), one could argue more abstractly that it is the ‘deviations from normal’ (texture contrast is a candidate

property in this case; see below), which draws attention, without a direct reference to any low-level feature. Our choice of using different categories of images helps to choose between these two alternatives. If the former hypothesis would be true, the attention drawing effects of LC modifications should hold for any image. But we fail to find any effect for *fractals*. This might be surprising for a bottom-up account, especially in the light of the previous finding that in the case of fractal images, low-level features play a stronger role (Parkhurst et al, 2002). The second hypothesis, on the other hand, already predicts this finding. Human subjects of course have prior knowledge (implicit or explicit) about natural scenes like forests, urban settings, or human faces. It follows that there is a certain expectation about what to see in such categories of images, in other words we know what is ‘normal’ for such scenes. Given modifications of high magnitude our subjects were confronted with parts of the image, which deviated from the normal for such images, hence their attention was drawn to those points. For fractal stimuli on the other hand, there is no such normal. A fractal image is a novel stimulus to inspect regardless of the local contrast modifications we introduced. Indeed informal interviews with subjects after the experiment indicated that the subjects were not aware of any modifications on fractal images, but noticed them to some degree in images from other categories. But there remains one more alternative explanation to exclude. One could argue that this normal is defined by LC only, and not by more abstract features as we posit. If the LC modifications were not enough to leave this normal range of LC in the case of fractals, this could make the subjects continue to investigate the image as an unmodified stimulus. Even though classical bottom-up models of saliency see the relationship between low-level features and attention linear (Itti and Koch, 2000), without any effective range, such reasoning would still be an attempt to attribute LC a causal role in attention. One finding speaks against this possibility. The fractal images and urban setting scenes used in the study had fairly similar second-order statistics. Therefore there is no reason to believe that modifications had different consequences for the LC profile of the two categories. Hence we conclude that positive modifications attracted attention, since they disrupted top-down expectations of subjects

5.4 *Negative LC modifications*

Now we turn to the effects of negative modifications. For *manmades*, *fractals*, and *faces* decreases in contrast had virtually no effect on fixation behavior. In a scenario, where LC has a causal role for the deployment attention, one would expect the avoidance of LC decreased regions, which is not confirmed by our findings. In the case of *naturals*, a slight decrease in LC (-20% peak modification) indeed repelled attention, supporting the causal role of LC. But a major decrease (-60% peak modification) had the reverse effect of attracting attention. Like in the case of LC increased images, there were more modifications on actual points than on control points. This result confirms the original reports, where this contrast modification paradigm was introduced (Einhäuser, and König, 2003; Einhauser et. al., 2006). Before we discuss why negative modifications can be attractive for visual system, we want to speculate on the effect differences between categories. Our urban setting and fractal images had higher luminance and texture contrasts. This leaves open the possibility that our major negative modifications were still in the range of contrast fluctuations found in such images. Even though this does not explain the negative finding for human face pictures, given that faces are processed differently than other natural images starting early in development (Hainline, 1978), this is still a plausible explanation. But even if this holds true, since negative modifications decrease LC to a substantial degree, we conclude that LC does not causally attract or repel attention during inspection of natural scenes, although it is correlated to attention.

5.5 *When does texture contrast guide eye movements?*

Can unintended texture contrast (TC) changes explain the results from the modified images? After Einhäuser and König (2003) argued that LC is not causally related to attention, and that top-down processes guide the fixation behavior, Parkhurst and Niebur (2004) developed a computational model taking into account TC as well as LC to simulate results on modified images. Regardless of how high in the visual system hierarchy TC is coded, it is still an intrinsic property of the image. If TC can explain the results, an attribution to top-down processes is inappropriate. Indeed their model was able to predict the fixation behavior of the subjects, since TC was increased both in contrast increased and decreased images. In our study, in unmodified images TC was slightly elevated for some categories at actual points compared to control points, showing that this second-order contrast can be useful for guidance of attention as well. In the case

of modified images the increase in TC at fixated points correlated very well with the saliency of our modifications for natural and urban environment scenes. In natural environment pictures in particular, we found an elevation of TC at actual points in negatively LC modified images, confirming the results of Parkhurst and Niebur (2004). But the modifications introduced similar TC changes (an increase in both negatively and positively modified images) in fractal images as well, where the modifications failed to change the fixation behavior of subjects. This provides support for Einhauser and colleagues' argument, that the description provided by Parkhurst and Niebur cannot exclude the possibility that LC and TC changes may be correlated with some higher-level property, rather than being directly linked to fixation behavior. If this high level property is deviation from normal, it is not hard to see why TC change explains the data from natural images, and not from fractal images. We believe that fixation behavior is at least partly under the control of such high-level properties, which are scene content related, and can in some circumstances be correlated with low-level features like LC or TC.

5.6 What guides the deployment of overt attention during viewing of natural images?

We see our results as providing evidence for the simultaneous operation of bottom-up and top-down mechanisms, which, as we will argue, are not mutually exclusive. Low-level features like LC and TC can in some situations be useful shortcuts to define the regions of interest in a given scene, which can explain why LC and TC are found to be high at fixation points. These linear mechanisms can be active during viewing of modified natural environment images, compatible with the finding that modifications at fixations increase gradually as the level of positive modification increases. But this mechanism is not enough, because it is not able to explain category-dependent effects observed in this study. Hence we argue that top-down processes, which make use of scene content, accompany the bottom-up process in natural image processing. An interesting account, which combines these two processes, comes from VanRullen (2003). He argues correctly that classical bottom-up models of saliency maps cannot explain the performance of humans and monkeys in rapid categorization tasks (e.g. Rousset et al, 2003; Mace et al., 2005). In these tasks subjects are shown natural scenes very briefly (around 30 ms), and in some cases in an unattended region of the visual field (Li et al., 2001; Fize et al., 2005). When asked to indicate if there was a target in this scene (animals, faces or means of transport) their accuracy is

as high as 90% and reaction times are as low as 150 ms. This poses the question of how a saliency map aimed at defining regions of interest on the basis of low-level features to deploy attention sequentially could be used in this context. The author argues that a definition of salience based only on the low-level features is not enough to account for such results. He extends the definition such that it includes together with various low-level contrasts, task instructions and behavioral relevance as well. In the present study, for instance, both high luminance (or any other) contrast regions and regions which do not fit overall scene-content become attractive for the visual system. An indirect evidence for the account comes from a large of body studies showing that the gist of the scene is registered before the details (for a review, see Hochstein and Ahissar, 2002), and this could explain the results from rapid categorization tasks. Even though it is clear that further experimental support is needed, a saliency definition combining top-down and bottom-up mechanisms can explain the results of our study.

5.7 The neural basis of Saliency Map

A comprehensive account of overt visual attention should not only explain why certain parts of a scene are more attractive than others, but should also be able to locate the mechanism in the brain. Candidate brain regions encoding saliency are together with others superior colliculus (Ignashchenkova et al., 2004), pulvinar (Posner and Petersen, 1990), frontal eye field (Thompson and Schall, 2000), none of which, to our knowledge, displays luminance contrast dependent firing. Two recent studies investigating brain regions involved in saliency encoding are of special interest for our study. Using parts of natural scenes Mazer and Gallant (2003) showed that half of the monkey V4 neurons they recorded displayed saliency-encoding behavior. The neurons with receptive fields as the targets of the next saccades were increasing their firing rate before the onset of the saccade. Additionally the activity of some of these neurons was changing according to the target identity, suggesting a top-down modulation. Gardner and colleagues recorded fMRI signals from humans as they viewed simple grating stimuli, which underwent contrast modifications. After adaptation to a certain contrast level V1, V2 and V3 activity increased with increasing contrast and decreased with decreasing contrast. The activity in V4, on the other hand, increased regardless of the direction of the change. This suggests that V4 encodes for salient changes in the stimuli. If this temporal change detection is true for changes in static natural

images as well, V4 would be a good candidate for guiding eye movements to our modified regions, regardless of the modification direction. Further research is needed to understand the relationship between different regions found to encode saliency and other regions involved in motor planning.

6. Conclusion

We confirm earlier findings (Einhauser and Konig, 2003; Einhauser et al., 2006) showing a correlative but not causal relationship between luminance contrast and overt attention. Additionally with the use of different image categories, we specify when texture contrast, a second-order contrast, can guide the eye movements, and when not. We conclude that during free viewing of natural images the overt attention is guided by interplay of bottom-up and top-down mechanisms.

Acknowledgements:

The author is grateful to Peter König and Selim Onat for their perfect guidance and support, to Wolfgang Einhäuser for careful discussions of the study, to Adriana Olmos for the pictures from ‘McGill Color Calibrated Image Database’, to Jonathan Camp for permission to use the fractal images, to Christian Kaul for the improved version of the modification code, to Wiebke Trost for data acquisition, to Sonja Engmann for the development of corrected KS-test.

7. References

Davon, D. (1977). Forest before the trees: the precedence of global features in visual perception. *Cognitive Psychology*, 9, 953 – 983.

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

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

Fize, D., Fabre-Thorpe, M., Richard, G., Doyon, B., and Thorpe, S. J. (2005). Rapid categorization of foveal and extrafoveal natural images: associated ERPs and effects of lateralization. *Brain & Cognition*, 59, 145 – 158.

Gardner, J. L., Sun, P., Waggoner, R. A., Ueno, K., Tanaka, K., and Cheng, K. (2005). Contrast adaptation and representation in the human early visual cortex. *Neuron*, 47, 607 – 620.

Hainline, L. (1978). Developmental changes in visual scanning of face and nonface patterns by infants. *Journal of Experimental Child Psychology*, 25, 90 – 115.

Hayhoe, M., and Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9, 188 – 194.

Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791 – 804.

Ignashchenkova, A., Dicke, P. W., Haarmeier, T., and Theier, P. (2004). Neuron – specific contribution the superior colliculus to overt and covert shifts of attention. *Nature Neuroscience*, 7, 56 – 64.

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

Johnson, A. P., and Baker, C. L. Jr. (2004). First- and second-order information in natural images: a filter-based approach to image statistics. *Journal of the Optical Society of America A*, 21, 913 – 925.

Kayser, C., Einhäuser, W., and König, P. (2003). Temporal correlations of orientations in natural scenes. *Neurocomputing*, 52, 117 – 123.

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

Krieger, G., Rentschler, I., Hauske, G., Schill, K., and Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spatial Vision*,

13, 201-214.

Li, F., VanRullen, R., Koch, C., and Perona, P. (2001). Rapid natural scene categorization without attention. *Paper presented at the Society of Neuroscience, San Diego, CA.*

Mace, M. J.-M., Thorpe, S. J., and Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: how robust at very low contrasts? *European Journal of Neuroscience, 21*, 2007 – 2018.

Mazer, A. M., and J. L., Gallant. (2003). Goal-related activity in V4 during free viewing visual search: evidence for a ventral stream visual salience map. *Neuron, 40*, 1241 – 1250.

Motter, B. C., and Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research, 38*, 1805 – 1815.

Navalpakkam, V., and Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*, 205-231.

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

Parkhurst, D.J., and Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16*, 125-154.

Parkhurst, D.J., and Niebur, E. (2004). Texture contrast attracts overt attention in natural scenes. *European Journal of Neuroscience*, *19*, 783-789.

Posner, M. I., and Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25 – 42.

Reinagel, P., and Zador, A.M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, *10*, 1-10.

Shofield, A. J. (2000). What does second-order vision see in an image? *Perception*, *29*, 1071 – 1086.

Tatler, B.W., Baddeley, R.J., and Gilchrist, I.D. (2005). Visual correlates of fixation selection: effects of time and scale. *Vision Research*, *45*, 643-659.

Thompson, K. G., and Schall, J. D. (2000). Antecedents and correlates of visual detection and awareness in macaque prefrontal cortex. *Vision Research*, *40*, 1523 – 1538.

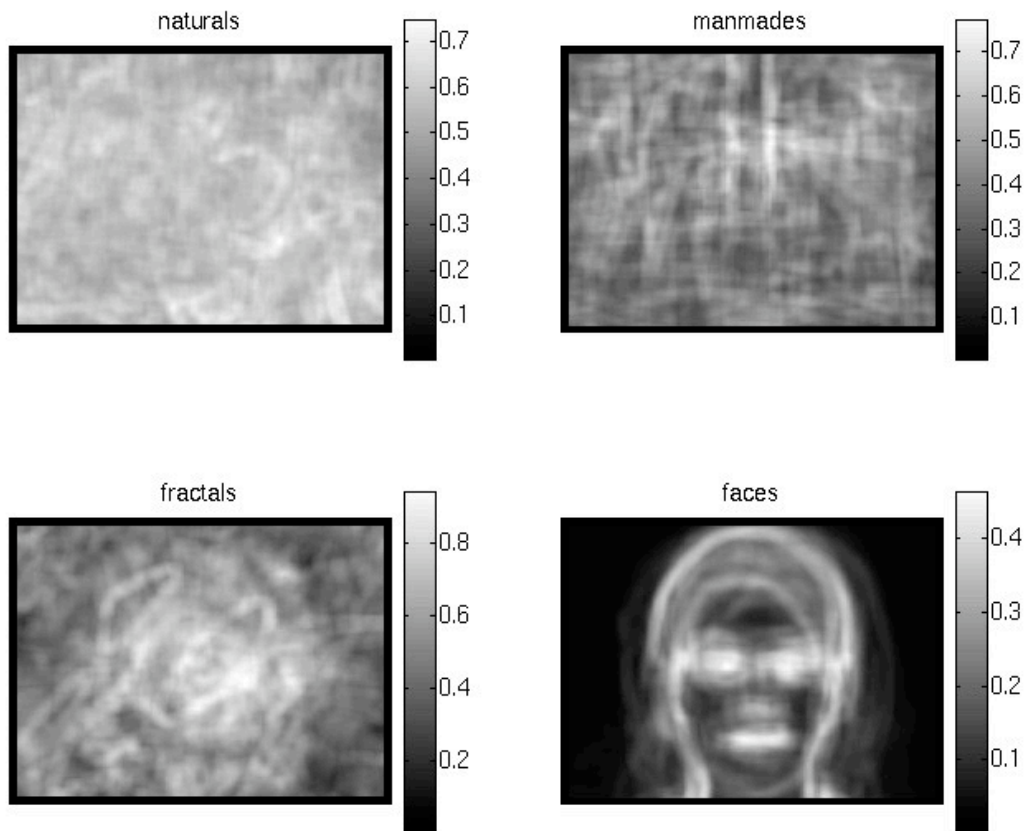
Torralba, A., and Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*, 391-412.

VanRullen, R. (2003). Visual saliency and spike timing in the ventral visual pathway. *Journal of Physiology Paris*, 97, 365 – 377.

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

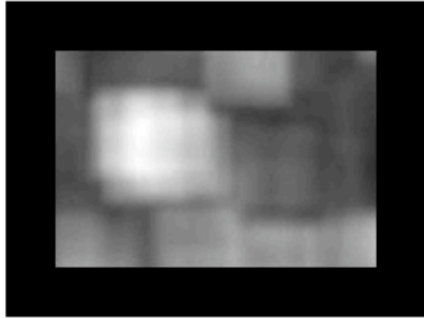
Supplementary Figures:

In the main text body, instead of describing every single step of our project, we tried to follow a clear story line, more or less like a scientific communication. Therefore in this section we provide the reader with additional figures for the sake of completeness.

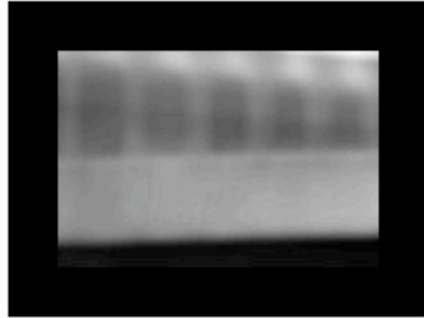


Supplementary Figure 1: The mean LC maps for each category. This figure shows that unlike many other image databases (e.g. Tatler et al., 2005) in our natural and urban environment pictures the high LC regions are not in the middle. The LC displays a rather uniform distribution throughout horizontal and vertical axes. The black borders show the regions where contrast cannot be calculated faithfully, due to the size of the patch, in which LC is computed.

naturals



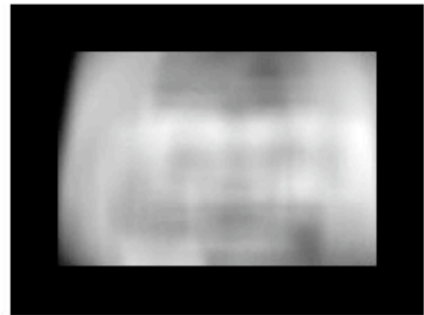
manmades



fractals



faces



Supplementary Figure 3: Example TC maps from each category. It is obvious that some structure image is preserved in these maps.



Supplementary Figure 2: The effect of modifications in natural images, once subjects separate, pictures collapsed (upper part), once pictures separate subjects collapsed. This figure shows that the unexpected repulsion effect at -20% modification level indeed is reliable.

Alper Açık
Arbeitsgruppe Neurobiopsychologie
Universität Osnabrück
Albrechtsraße 28
49069 Osnabrück
Tel.: 0541/9693509
email: aacik@uos.de

Aufklärung / Einwilligung

Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

Sie haben sich freiwillig zur Teilnahme an dieser Studie gemeldet. Hier erhalten Sie nun einige Informationen zu Ihren Rechten und zum Ablauf des folgenden Experiments. Bitte lesen Sie sich die folgenden Abschnitte sorgfältig durch.

1) Zweck der Studie

Ziel dieser Studie ist es, neue Erkenntnisse über den Einfluss von Kontrast auf die Wahrnehmung von natürlichen Bildern zu erhalten.

2) Ablauf der Studie

In dieser Studie werden Ihnen in 4 Blocken jeweils 144 Bilder auf einem Computermonitor gezeigt. Bitte sehen Sie sich die Bilder aufmerksam an

Um Ihre Blickposition zu errechnen, wird Ihnen ein "Eye-Tracker" auf den Kopf geschnallt. Dieses Gerät erfasst die Position Ihres Auges mit Hilfe von kleinen Kameras und Infrarotsensoren. Dieses Verfahren ist ein psychometrisches Standardverfahren, das in dieser Art bereits vielfach angewandt und getestet wurde. Bei unseren bisherigen Erfahrungen und Experimenten mit dem Gerät ist keine Versuchsperson zu Schaden gekommen.

Zu Beginn der Untersuchung muss der "Eye-Tracker" eingestellt werden, dieser Vorgang dauert etwa 5-15 Minuten. Das eigentliche Experiment dauert dann etwa 30 Minuten. Der Versuchsleiter wird während des ganzen Experiments mit Ihnen im Versuchsraum sein und steht Ihnen für Fragen jederzeit zur Verfügung. Nach der Studie erhalten Sie weitere Informationen zum Sinn und Zweck dieser Untersuchung. Bitte geben Sie diese Informationen an niemanden weiter um die Objektivität eventueller Versuchspersonen zu wahren.

3) Risiken und Nebenwirkungen

Diese Studie ist nach dem derzeitigen Wissenstand des Versuchsleiters ungefährlich und für die Teilnehmer schmerzfrei. Durch Ihre Teilnahme an dieser Studie setzen Sie sich keinen besonderen Risiken aus und es sind keine Nebenwirkungen bekannt. Da diese Studie in

ihrer Gesamtheit neu ist, kann das Auftreten von noch unbekanntem Nebenwirkungen allerdings nicht ausgeschlossen werden.

Wichtig: Bitte informieren Sie den Versuchsleiter umgehend, wenn Sie unter Krankheiten leiden oder sich derzeit in medizinischer Behandlung befinden. Teilen Sie dem Versuchsleiter bitte umgehend mit, falls Sie schon einmal einen epileptischen Anfall hatten. Bei Fragen hierzu wenden Sie sich bitte an den Versuchsleiter.

4) Abbruch des Experiments

Sie haben das Recht, diese Studie zu jedem Zeitpunkt und ohne Angabe einer Begründung abbrechen. Ihre Teilnahme ist vollkommen freiwillig und ohne Verpflichtungen. Es entstehen Ihnen keine Nachteile durch einen Abbruch der Untersuchung.

Während des Experimentes haben Sie einmal die Gelegenheit zu einer Pause, in dieser Zeit kann Ihnen auch der "Eye-Tracker" abgenommen werden. Auch falls Sie eine weitere Pause wünschen oder auf die Toilette müssen, ist dies jederzeit möglich.

Sollten Sie zu irgendeinem Zeitpunkt während des Experiments Kopfschmerzen oder Unwohlsein anderer Art verspüren, dann informieren Sie bitte umgehend den Versuchsleiter.

5) Vertraulichkeit

Die Bestimmungen des Datenschutzes werden eingehalten. Personenbezogene Daten werden von uns nicht an Dritte weitergegeben. Die von Ihnen erfassten Daten werden von uns anonymisiert und nur in dieser Form weiterverarbeitet oder veröffentlicht.

6) Einverständniserklärung

Bitte bestätigen Sie durch Ihre Unterschrift die folgende Aussage:

“Hiermit bestätige ich, dass ich durch den Versuchsleiter dieser Studie über die oben genannten Punkte aufgeklärt und informiert worden bin. Ich habe diese Erklärung gelesen und verstanden. Ich stimme jedem der Punkte zu. Ich ermächtige hiermit die von mir in dieser Untersuchung erworbenen Daten zu wissenschaftlichen Zwecken zu analysieren und in wissenschaftlichen Arbeiten anonymisiert zu veröffentlichen.

Ich wurde über meine Rechte als Versuchsperson informiert und erkläre mich zu der freiwilligen Teilnahme an dieser Studie bereit.”

.....
Ort, Datum

.....
Unterschrift

.....
Bei Minderjährigen, Unterschrift des Erziehungsberechtigten