

## Accepted Manuscript

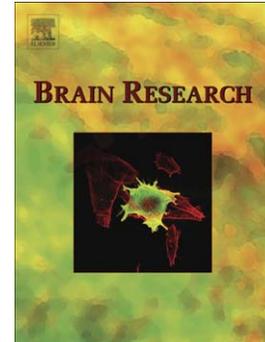
Viewing strategy of *Cebus* monkeys during free exploration of natural images

Denise Berger, Antonio Pazienti, Francisco J. Flores, Martin Nawrot,  
Pedro E. Maldonado, Sonja Grün

PII: S0006-8993(11)01873-7  
DOI: doi: [10.1016/j.brainres.2011.10.013](https://doi.org/10.1016/j.brainres.2011.10.013)  
Reference: BRES 41840

To appear in: *Brain Research*

Accepted date: 7 October 2011



Please cite this article as: Denise Berger, Antonio Pazienti, Francisco J. Flores, Martin Nawrot, Pedro E. Maldonado, Sonja Grün, Viewing strategy of *Cebus* monkeys during free exploration of natural images, *Brain Research* (2011), doi: [10.1016/j.brainres.2011.10.013](https://doi.org/10.1016/j.brainres.2011.10.013)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Viewing strategy of *Cebus* monkeys during free exploration of natural images

## **Denise Berger<sup>1</sup>**

Neuroinformatics, Inst. Biology, Freie Universität, Königin-Luise Str. 1-3, Berlin, Germany  
Bernstein Center for Computational Neuroscience, Phillipstr. 13, Berlin, Germany  
d\_berger@gmx.net

## **Antonio Pazièti**

European Brain Research Institute  
Via del fosso di Fiorano 64, Rome, Italy  
apazièti@gmx.net

## **Francisco J. Flores<sup>2</sup>**

Program of Physiology and Biophysics, Faculty of Medicine, University of Chile,  
Av. Independencia 1027, Santiago, Chile  
fjflores@mit.edu

## **Martin Nawrot**

Neuroinformatics, Inst. Biology, Freie Universität, Königin-Luise Str. 1-3, Berlin, Germany  
Bernstein Center for Computational Neuroscience, Phillipstr. 13, Berlin, Germany  
nawrot@neurobiologie.fu-berlin.de

## **Pedro E. Maldonado**

Program of Physiology and Biophysics, Faculty of Medicine, University of Chile,  
Av. Independencia 1027, Santiago, Chile  
pedro@neuro.med.uchile.cl

## **Sonja Grün**

Institute of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Research  
Center Jülich, 52425 Jülich, Germany  
Germany Theoretical Systems Neurobiology, RWTH Aachen University, 52056 Aachen, Germany

s.gruen@fz-juelich.de

\* Corresponding author:

Dr. Denise Berger, Laboratory of Neuromotor Physiology, IRCCS Fondazione Santa Lucia, Via  
Ardeatina 306, 00179 Roma, Italy, d\_berger@gmx.net

<sup>1</sup> *Present address:* Laboratory of Neuromotor Physiology, IRCCS Fondazione Santa Lucia, Via  
Ardeatina 306, 00179 Roma, Italy,

<sup>2</sup>*Present address:* Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, USA; and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, USA.

## **Abstract**

Humans and other primates move their eyes several times per second to foveate at different locations of a visual scene. What features of a scene guide eye movements in natural vision? We recorded eye movements of three monkeys during free exploration of natural scenes and propose a simple model to explain their dynamics. We use the spatial clustering of fixation positions to define the monkeys' subjective regions-of-interest (ROI) in natural scenes. For most images the subjective ROIs match significantly the computed saliency of the natural scene, except when the image contains human or primate faces. We also investigated the temporal sequence of eye movements by computing the probability that a fixation will be made inside or outside of the ROI, given the current fixation position. We fitted a Markov chain model to the sequence of fixation positions, and find that fixations made inside a ROI are more likely to be followed by another fixation in the same ROI. This is true, independently of the image saliency in the area of the ROI. Our results show that certain regions in a natural scene are explored locally before directing the focus to another local region. This strategy could allow for quick integration of the visual features that constitute an object, and efficient segmentation of objects from other objects and the background during free viewing of natural scenes.

# 1. Introduction

Early studies by Stratton (1902; 1906) showed that free exploration of natural scenes is performed through a spatiotemporal sequence of saccadic eye movements and ocular fixations. This sequence indicates the focus of spatial attention (Noton and Stark, 1971a; Biedermann, 1987; Crick and Koch, 1998), and is guided by bottom-up and top-down attentional factors. Bottom-up factors are related to low-level features of the objects present in the scene being explored (Treisman and Gelade, 1980; Koch and Ullman, 1985; Itti and Koch, 1999, 2001) while top-down factors depend on the task being executed while exploring a scene (Buswell 1935; Yarbus, 1967; Just and Carpenter, 1967), the context in which those objects are located (Torralba, et al. 2006), and the behavioral meaning of the objects being observed (Guo et al., 2003; Guo et al., 2006). For example, traffic lights can attract attention and eye movements both by bottom-up and top-down factors: they are very salient in virtue of their low-level, intrinsic properties (color and intensity), and also very meaningful to the driver (behavior and context).

Several computational models have been proposed to explain guidance of eye movements and attentional shifts during free viewing of natural scenes (e.g., Wolfe, 1994; Milanese et al., 1995; Tsotsos et al., 1995; Itti et al. 1998). The most common strategy includes the computation of saliency maps to account for bottom-up factors and that define the regions-of-interest (ROIs) that attract eye movements. The saliency maps are then fed into a winner-take-all algorithm to account for the top-down attentional contribution (Milanese et al., 1995; Itti et al., 1998). During the execution of specific visual search tasks, the nature of the task itself can be used to estimate contextual, task-relevant scene information that will add up to the saliency model (Torralba et al 2006). However, during free viewing of natural scenes, where no particular task is executed, it is more difficult to estimate the appropriate context.

Furthermore, although meaningful objects populate natural scenes, there are currently no computational tools that allow to link behaviorally relevant images and exploration strategies solely based on local or global features.

We hypothesize that the spatial clustering of ocular fixations provides a direct indication of the subjective ROIs in a natural scene during free viewing conditions. It is very likely that subjective ROIs include both top-down and bottom-up attentional factors, thereby potentially providing a framework to formally understand the guidance of eye movements and spatial attention by studying the transitions between and within regions. The approach presented here provides several advantages: first, the use of monkeys additionally allows the recording of single neurons (cmp. Maldonado et al., 2008). Second, it presents a tool to classify fixations that enables to relate neuronal activity to natural behavior (see Discussion), without making assumptions about the meaning of the images to the observer. Third, our approach can be generalized to eye movements of humans. We find that in most cases, the subjective ROIs match well both the objects in the scene and the ROIs defined by their saliency maps. Exceptions are scenes containing human or primate faces.

We made use of a Markov chain (MC) analysis to investigate the sequences of visited ROIs (assumed to be the *states* of a random walk) and extract their probabilities. Our approach of the scanpath analysis differs from Feng (2006) (reading task experiment), Van Der Lans et al. (2008) (search task), and Simola et al. (2008) (word search task) in that we feed the MC algorithm with the extracted ROIs. Such an investigation of fixation sequences shows that during free viewing of natural scenes a fixation is most likely to occur within the same ROI where the previous fixation occurred, suggesting that local object exploration is executed before directing the focus to a new ROI.

## 2. Results

### *2.1. Fixations on natural images are spatially distributed in clusters and define subjective regions-of-interest*

Three monkeys (D, M, and S) participated in an electro-physiological experiment over many sessions, in which they were exposed to different natural images for 3-5 sec, interleaved with blank screens or blank screens with a fixation spot (see Figure 1, and Section 4.1 for details). Their eye movements were recorded with a scleral search coil, while the animals were allowed to freely explore the monitor screen with self-initiated eye movements (see Figure 2A as an example of one image overlaid by an exemplary scanpath and the respective fixations).

An automatic algorithm extracted the fixations and saccades performed by the monkeys from the vertical and horizontal eye movements (Figure 2B, see Section 4.2. for details), and derived the distributions of fixation and saccade durations (Figure 2A, B). The distributions of fixation durations derived from all sessions and for all images (Figure 2C) of monkey D and M have very similar shapes, and their mean fixation durations are 310 msec and 240 msec, respectively. These values correspond well to average fixation durations reported for humans during exploration of natural scenes, found to be in the range between 260 and 330 msec (Castelhano, 2007; Ossandon et al., 2010). However, the distribution of fixation durations of monkey S (Figure 2C, red) differs from the distributions of the two other monkeys: it is broader, less skewed and has a heavy tail, and exhibits a much longer mean fixation duration (420 msec). Interestingly, the distributions of saccade durations of the three monkeys (Figure 2D) differ only little (mean saccade durations: 32.1 msec, 31.0 msec, and, 33.8 msec for monkeys D, M, and S, respectively).

In a next step we investigated how the eye movements of the three monkeys were spatially distributed on the viewed images, and if these also show differences between monkeys D and M, and S.

The spatial distribution on one specific image was derived from eye movements across all presentations of the image. We observed that the spatial distributions of fixations of monkeys D and M exhibit dense spatial clusters that are related to conspicuous objects in the underlying images (see examples for four different images in Figure 3). The positions of the clusters are qualitatively similar for both monkeys for the same image, but are qualitatively different for each individual image (Figure 3, columns 1, 2).

However, the spatial fixation distributions of monkey S are unique: more than 90% of his fixations are evenly distributed inside a large cluster in the lower left quadrant of the images. This pattern is conserved across different images, and seems independent of the content of the images (Figure 3, column 3), indicating that the eye movements of this monkey were not related to the images. It is unlikely that the differences in fixation duration and of the exploration patterns of monkey S were due to a physiological dysfunction of his oculomotor system, since his saccade durations were very much in agreement with the other monkeys (Figure 2D), indicating an intact saccade generation mechanism. Inspection of the fixations on images containing only a fixation spot, routinely presented just before each natural image to detect potential artifacts and eye calibration issues, shows that monkey S did fixate on the fixation spot within the required limits. Therefore we concluded that the monkey adopted an unusual strategy to get rewarded, deliberately gazing over the images without paying attention to the images contents. We include data from this animal both as a comparison to the other monkeys, and as a potential methodological issue for further studies.

For monkeys D and M, we assume that each of the spatial fixation clusters represents a subjective ROI. The position of subjective ROIs on an individual image is likely to depend on at least two factors: a bottom-up image feature driven component and a top-down attentional factor. To explore the contribution of the bottom-up component on the spatial positions of the subjective ROIs, we compare in a next step the similarity of the map of the fixations with the saliency map of the respective image.

## 2.2. Fixation and saliency maps diverge for images containing faces

We computed the saliency maps of the images based on the model described by Walther and Koch (2006) (see examples in Figure 4A). Simultaneously we computed the fixation maps for each image and monkey by down-sampling the original 800 x 600 pixels-images to 30 x 40 pixels-images and normalizing the original fixation distribution (details in Section 4.4). The similarity of each fixation map and its corresponding saliency map was derived by computing the symmetrized Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951) between the two distributions (Rajashekar et al., 2004, details in Section 4.5). The smaller the KLD, the higher the similarity between the two distributions, with its lower bound at zero, if the two are identical. To evaluate the significance of  $KLD_{act}$  of the actual, measured data, we calculated the probability distribution of  $KLD_{ind}$  values derived from the same saliency map but with fixation maps resulting from a random viewer, i.e., randomly (homogeneously) distributed fixation points on the image (Parkhurst et al. 2002, for details see Section 4.5). This procedure implies the assumption of independence between the two maps, and allowed us to test if the monkeys' viewing behavior deviates significantly from a non saliency-related behavior (Figure 4A, B).

The results for all monkeys and all images are shown in Figure 4C. For visualization purposes we show for each image the difference of the actual  $KLD_{act}$  and the mean  $\langle KLD_{ind} \rangle$  of the  $KLD_{ind}$  - distribution,  $\Delta KLD = \langle KLD_{ind} \rangle - KLD_{act}$  (color bars in Figure 4C). In 8 out of 11 images explored by monkey D (Figure 4C, blue bars) we find significant positive  $\Delta KLD$  values (i.e.,  $KLD_{act} \ll \langle KLD_{ind} \rangle$ ) ( $p < 0.01$ , marked by asterisks), and similarly for monkey M (significant: 3 out of 4 images; Figure 4C, green bars), indicating that for monkeys D and M the saliency maps of these images were good predictors of the fixation positions. However, in the remaining 25% of images, the  $\Delta KLD$  was

significantly negative (i.e.,  $KLD_{act} \gg KLD_{ind}$ ) when compared to a random viewer, i.e., the fixation map differs significantly ( $p > 0.01$ ) from the saliency map, leading to the conclusions that here a) the saliency maps were not predictors of the fixation positions, and b) the viewing behavior differed from random viewing, indicating the presence of a distinct viewing strategy for these images. Interestingly, this holds true for all images that differ in content from the other images in that they show *faces* of human or non-human primates, and not for the other images, which contained only non-primate animals. Performing the same analysis only on fixations that belonged to ROIs did not alter the significance of our results (cmp. Methods, Section 4.5).

The analysis of the previous section already hinted at differences of the viewing behavior of monkey S as compared to monkeys D and M. Our quantitative analysis of the similarities of the saliency and fixation maps additionally showed marked differences between monkey S to the other two monkeys: the fixation patterns of monkey S never deviates significantly from a random viewer (Figure 4C, brown bars), thus confirming our hypothesis that this monkey did not actively explore the images. In fact, it seems that he just kept his gaze within the lower left part of the screen, independently of the presented image (Figure 3, column 3).

In summary, we find that low-level image features drove the fixations performed by the monkeys that actively explore the natural scenes if the images did *not* show faces of primates. For the remaining images, most of the eye movements relate to the faces, regions that typically are of low saliency value and thus have a low bottom-up impact.

### *2.3. Consecutive fixations are likely to stay inside of the same fixation cluster.*

Our analysis of the fixation positions (Section 2.1) revealed that these are not evenly distributed across the images, but rather tend to occur clustered in space (Figure 3). Our interpretation was that

these clusters represent subjective ROIs. Thus, our next aim is to gain insight on the temporal sequences of visiting the subjective ROIs. Therefore we explored the scanpaths of the image explorations by applying a Markov chain (MC) analysis to the eye movement trajectories (see details on Section 4.5). Thereby we assume each of the significantly identified fixation clusters on a particular image as a Markov state, and estimate the probabilities for consecutive fixations to either stay in the same cluster, to switch to a different cluster, or to land into the background. In this analysis the assumption of a MC enters in that the next state will be reached only depending on the current state, but does not depend on past states (see details on Section 4.5).

The cluster analysis of the fixation positions typically revealed 3 to 5 significant clusters per image for monkeys D and M, however, not a single significant cluster could be extracted for monkey S. Thus this monkey seems not to express subjective ROIs, and we had to conclude that this monkey is not actively exploring the images. Since the MC analysis is based on ROIs, monkey S had to be excluded from the subsequent analysis of the sequence of fixation positions.

Figure 5A shows examples of eye movement sequences (4 out of 33) of monkey D during presentations of the same image. The fixation positions of monkey D on the image during all its presentations were grouped into three significant clusters (Figure 5B, color coded). Fixation positions that do not belong to any identified cluster (small blue dots) are pooled together and assigned to the background cluster (see Section 4.6 & 4.7). The result of the MC analysis on these data is shown in Figure 5C as a transition graph. Each identified significant cluster, as well as the background cluster, represents a state of the model, whereas the transitions between the states (whose probabilities are indicated in black) are marked by directed arrows. The statistical significance was evaluated by comparing the transition probabilities of the empirical data to uniform probabilities (Figure 5C, numbers in gray; details see Section 4.7). The probabilities (across all images) of staying within the significant clusters are 87% (40/46) for monkey D and 95% (19/20) for monkey M, thus significantly

higher than expected by chance (Figure 5D). In contrast, the probabilities of moving between significant clusters (Figure 5E) are significantly lower than chance (in 62% of the transitions for monkey D (60/97) and in 62% for monkey M (83/134)). (Note, that for the latter analyses the background state was discarded.) This also holds true separately for monkeys D and M (Wilcoxon test,  $p < 0.001$ ).

As for saccade durations (Figure 2D), the distribution of saccade lengths is skewed, thus showing a tendency for shorter with respect to longer saccades. In order to avoid any bias due to the skewness of the distribution, we performed a second test, which, instead of uniform probabilities, takes into account the actual saccade amplitude distribution. The expected transitions were weighted by the actual probability of saccade amplitudes (see Methods, Sec. 4.7 for details). The results confirmed the previous analysis, i.e., significant larger probability of staying within a cluster and significant lower probability of switching between clusters than expected (Figure 5D, E in green).

Overall, the Markov chain analysis revealed that the monkeys preferentially move their eyes within the same ROI before saccading out to another ROI or to the background. These results did not show any dependence on the contents of the images, in particular with respect to primate faces. Thus, the viewing strategy of the monkeys seems to be composed of sequences of local explorations of regions-of-interest, but not of random eye movements between ROIs.

### **3. Discussion**

The present work shows that during free viewing of natural images, *Cebus* monkeys follow a strategy that involves periods of local exploration, characterized by consecutive fixations that stay

inside the same regions-of-interest. These periods of local exploration are typically followed by longer saccades into a new ROI, where a new period of local exploration begins. ROIs were defined as areas containing clusters of fixations performed by the monkeys over several presentations of an image. For most of the images, the locations of the fixation clusters correlate well with saliency maps, suggesting low-level features as the driving force for the eye movements. Images containing faces are an exception. In the latter images, faces attract most of the fixations despite their very low saliency. Therefore, as hypothesized, subjective ROIs reflect both bottom-up and top-down factors. Our approach based on subjective ROIs is robust with respect to content and semantic meaning of the images, because it relies on the spontaneous sequences of eye movements performed by the subject. Similar approaches have been used in humans, showing conserved clusters of fixations in the same image for different subjects (Judd et al., 2009).

Our analysis of eye movement sequences during free viewing is based on the finding that fixations are not evenly distributed on an image, but rather define clusters, on top of conspicuous objects. This was the case for two out of three subjects studied (monkeys D and M). However, the third monkey (S) used a completely different viewing strategy. His fixations were always restricted onto the lower left half of the images and were evenly distributed therein. No significant clusters could be extracted from his fixations, and did not show any significant correlation between fixation maps and saliency maps, which corresponds to a random viewing behavior. Given that the distributions of saccade durations of the three monkeys were undistinguishable (Figure 2D), we concluded that it is unlikely that this monkey had any deficiency in the oculomotor system. We rather assume that monkey S did not actively explore the images. Our experimental design could not prevent this to happen, because the monkeys were only required to keep their gaze within the limits of the screen to be rewarded. It is very likely, that the monkey did not only learn to keep his gaze within the limits of the screen, but additionally within a specific region therein while ignoring the images. Our explanation

relates to the process of training. During many weeks the monkeys needed to be trained to fixate on the central point. Initially the window to get a reward was large and was progressively downsized. Monkey S may have learned that natural images were no different than fixation images and that by trying to keep his gaze in some specific area of the screen, he will get a reward (which he did). This strategy enabled this animal to get rewarded only by trying to avoid moving the eyes far away from a particular region of the screen, hence the particular fixation distribution. Therefore we restricted our analysis to the scanpaths of the monkeys that explored the images, and we limit our discussion to the results we derived from monkeys D and M.

The visual fixations of monkeys D and M cluster on locations of the images that appear to be relevant to the monkeys, and thus we interpret these clusters as subjective ROIs. Similar viewing behavior had been found in humans that were freely exploring natural images: most of the fixations were made in the same regions of an image *across* observers. In fact, fixation locations from one observer can be used to predict the locations where other observers will fixate (Judd et al, 2009). Therefore, the images can be segmented into informative and redundant regions both for monkeys and humans (Mackworth and Morandi, 1967; Yarbus, 1967; Krieger et al., 2000). A common way to segment natural images is to apply saliency analyses. In our study we were interested in isolating the contribution of low-level features – such as orientation, color and intensity – and to relate it to the locations of the fixation clusters. In order to extract this relation we used the saliency model of Walther and Koch (2006). Saliency turned out to be a good predictor for the fixation positions. This suggests that during free viewing the eye movements are mainly driven by low-level features. Images containing primate faces are an exception (for monkeys and humans), in that the fixation positions cluster in regions containing no salient features. The mismatch between saliency and positions of fixation clusters can be attributed to the influence of top-down mechanisms, where attention to meaningful details of the

objects determines the location of gaze. This result fits well with data from human studies where the choice of fixation positions has been shown to be either driven by bottom-up (exogenous) or by top-down (endogenous) factors (Mackworth and Morandi, 1967; Cerf et al., 2008). It has also been shown that the saliency model does not account for fixations that were directed to the eyes of humans (Birmingham, 2009). Thereby, faces appear to play a particular role, being probably the most important visual stimuli in primate social communication (Bruce and Young, 1998), as they can provide significant cues to intention and mental state of other individuals (Bruce and Young, 1998; Emery, 2000; Andrew, 1963; Anderson, 1998). Similar observations were found in non-human primates: monkeys make longer fixations on faces (Guo et al., 2006), and respond appropriately to the expressions of other individuals (Mendelson et al., 1982), and are able to recognize their faces (Rosenfeld and Van Hoesen, 1979).

Psychological studies have shown that the sequences of saccades and fixations are relevant for perception (Noton and Stark, 1971b). In humans, during free viewing of still images for long time periods (i.e., > 10 sec) saccade amplitudes decrease exponentially (Antes, 1974; Unema et al., 2005). Pannasch et al. (2008) showed that fixation durations increase after the first two seconds of exploration, revealing a global image exploration that spans the first two seconds, followed by a local, feature exploration phase, evident after four seconds of exploration. The maximum exploration time in our study was five seconds, which could suggest that the higher probability of staying inside a cluster is a consequence of the late, local exploration phase. However, examination of the raw data (see for example Figures 2A and B, and Figure 5A) reveals that some consecutive fixations are separated by short saccades even during the first seconds of exploration.

We find that the monkeys fixate preferably at certain restricted locations on the images (identified as clusters of fixations), and that the eye movements between these clusters are not random. The Markov chain analysis revealed that the monkeys primarily make short saccades within a cluster of

fixations. These short saccades are likely to be followed by a larger saccade, that directs the gaze to a new position inside a different cluster. This finding is consistent with the hypothesis that large saccades to new areas are followed by local, short saccades to nearby positions for refinement of the percept (Körner et al., 1999; Ullman, 1995). Further studies showed that applying a Markov model to humans freely viewing advertisement has revealed similar local vs. global exploration modes (Wedel et al., 2008), and that humans and monkeys attend to relevant social stimuli when viewing short movies (Shepherd, 2010). Thus our results indicate that the monkeys use a similar strategy for scanning natural images as humans do.

Experiments including active vision, i.e., without the request that eyes fixate on a pre-defined position, are infrequently included in studies that involve electro-physiological recordings, as they do not contain repetitive, identical trials and thus are harder to analyze. This study provides new approaches to data from free viewing animals and thus opens new routes for experiments that aim to relate neuronal activities to natural behavior. The Markov chain model appears to be a natural way to compress complex and variable data sets such as eye movements made on natural images. Clusters can be labeled and further grouped into different categories by saliency analysis or image segmentation methods, and the eye movements can be represented as a Markov state graph, which assigns probabilities to the transitions between states (as shown in Figure 5). Such a procedure offers the possibility of summarizing an otherwise very disparate data set. Neurophysiological data could be subsequently analyzed in the context of the different categories of fixation clusters.

Electro-physiological studies that involve the presentation of natural stimuli, either during free viewing or fixed gaze, already show that the perspective of a simple stimulus-response relation explains only partially the neural activity observed in natural vision (Yen et al., 2008). In these situations, neuronal activity appears much more complex, which cannot be simply related to the stimulus features, where higher-order brain areas and attentional effects obviously play a crucial role. Active vision includes

self-initiated eye-movements and thus naturally involves a combination of internal and external driving forces. Active vision is fast: within the duration of a fixation (about 200 msec) visual input enters the system, visual information is processed and the next new eye movement is initiated. This requires fast processing and leaves to every individual stage of the nervous system only very limited time for computation (Thorpe et al., 1996). This limited time can be better used if some consecutive fixations are made close to each other, serially grouping object features (Houtkamp and Roelfsema, 2010).

Thus, electro-physiological studies of active vision need to include the dynamics of processing, as suggested by some of the models of the visual system (Van Rullen et al., 1998; Körner et al., 1999), which predict temporal coordination of neuronal activities. Recently, we found that spike synchrony is involved early in the processing in the visual system (Maldonado et al., 2008) and that a signal activating large populations of neurons of the visual system (local field potential) occurs locked to saccade onset, thereby providing an internal reference signal for the coordination of neuronal activity induced by visual input (Ito et al., 2011). It is highly likely that this signal is modulated along the scanpath or has an attentional function thus providing the ground for context-dependent neuronal processing.

## **4. Methods**

### *4.1 Animals and surgery*

All experiments followed the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were in accordance with University of Chile guidelines. All surgical and recording procedures are described in Maldonado et al., 2008. Three adult, male capuchin monkeys (*Cebus apella*) weighing 3-4 kg served as subjects for this study. Henceforth, these animals are referred

to as monkey D, M, and S. Under sterile conditions, each animal was implanted with a scleral search coil for monitoring eye position (2 kHz sampling rate, DNI Instruments, Resolution: 1.2 min of arc; for details see Judge et al., 1980), and a cranial post for head fixation. During the experiment, the animals were seated in a chamber dimly lit at a low scotopic level (1-2 lx, LX-110 Lux Meter). They were presented with a collection of 11 (monkeys D and S) and 4 (monkey M) pictures of different natural scenes (consisted of pictures of animals, faces and landscapes, 800 x 600 pixel resolution; taken from Corel® photo library). The pictures were displayed on a CRT computer monitor (frame rate: 60 Hz) located 57 cm in front of the animals, subtending  $40^\circ \times 30^\circ$  of visual angle. As a control, for every third stimulus presentation, a blank frame with black background was presented instead of a natural image. We refer to the trials with natural image stimuli as image condition trials and those with the blank frame as blank condition trials. In order to maintain the alertness of the animals, and to control eye coil precision, they were trained to perform a fixation task before every trial, in which a black frame with a single fixation spot was presented and they had to fixate it ( $1^\circ$  window) for 1 sec in order to be rewarded (referred to as fixation cue). Then, a natural image or the blank frame was presented for 3 or 5 sec for monkey D or S and M, respectively (free viewing trials) (Figure 1). In the free viewing trial, the animals were allowed to freely explore the monitor screen with self-initiated eye movements while the experimental protocol required the animals to maintain their gaze within the limits of the monitor for the whole presentation period, to be rewarded with a drop of juice. A session was composed of image condition trials and blank condition trials alternating with fixation cues. Before each session we calibrate the coil with a series of fixation cues, referred to as fixation epoch. If the monkeys were willing to continue to work after a session we ran a further session starting with a fixation epoch, followed by a new set of images. This process was repeated as long as the animals were motivated to continue the task. Only the data collected during the presentation of fixation cues and natural images served for the following analyses and defined an experimental session.

## *4.2 Detection and statistics of eye movements and fixations*

We developed an automatic algorithm to extract different types of eye movements from the eye traces (Maldonado et al., 2008). Eye movements were categorized in two different groups (saccades and fixations) (cf. Figure 2A, B), according to the following criteria: Saccades were defined as eye movements with an angular velocity higher than  $150^\circ/\text{sec}$  and lasting for at least 5 msec, and exhibit a minimum acceleration of  $170^\circ/\text{sec}^2$ . Fixation periods were defined as gaze positions lasting at least 100 msec within  $1^\circ$  of the gaze location, following a saccade. Data that could not be assigned into one of the two categories (e.g., drifts) were not taken into account for further analysis. Only pairs of unambiguous saccade-fixation (S-F) sequences were considered for further analysis. Basic statistics of fixation and saccade durations pooled per monkey over all sessions are shown in Figure 2C, D.

## *4.3 Computation of Saliency Maps*

In order to relate the visual foci of the monkeys as expressed by the fixation positions to the features of the images, we computed maps of fixation points ('fixation maps'; see Section 4.4) and separately, maps of salient features of the images ('saliency maps'), and correlated the two (cf. Section 4.5). A saliency map is a topographically arranged map that represents visual saliency of a corresponding visual scene. Koch and Ullman (1985) proposed to combine different visual features that contribute to attentive selection of a stimulus (e.g., color, orientation, movement etc.) into one single topographically oriented map (saliency map), which integrates the normalized information from individual feature maps into one global measure of conspicuity. We concentrated here on a saliency map model by Walther and Koch (2006) that ignores the motion aspect, but uses color, intensity, and

orientation (implementation freely available at <http://www.saliencytoolbox.net/>). Thereby, the images were segregated into three separate feature maps: one for intensity, one for color, and one for orientation. In a second step, each feature was re-organized into a center-surround arrangement characteristic of receptive field organization (Hubel and Wiesel, 1962), and highlights the parts of the scene that strongly differ from their surroundings. This was achieved by computing the differences between fine and coarse scales applied to the feature maps to extract locally enhanced intensities for each feature type. In the last step these resulting conspicuity maps were normalized to the total number of maps and added to yield the final saliency map  $s(x, y)$  (see examples in Figure 4A).

#### *4.4 Computation of Fixation Maps*

As a measure of the regions of the images that preferably attract the interest of the monkeys we computed a ‘fixation map’ for each image and monkey. All fixations performed by a monkey on a particular image were pooled across different sessions and trials (see examples in Figure 3A) to calculate a two-dimensional probability distribution of the fixations  $f(x, y)$ . Such a distribution was derived by discretizing the images into two dimensional bins of 30 x 40 pixels (original resolution was 600 x 800 pixel), counting the fixation positions in each of them and normalizing the counts to the total number of fixations on that image to yield a probability distribution. This resolution corresponds to approximately 1° of viewing angle in x- and y-dimension (1° corresponds to 1 cm on the screen which is located 57 cm in front of the monkey), which was also chosen as the tolerance for the definition of a fixation.

#### *4.5 Similarity of saliency and fixation maps*

To quantify the similarity between the saliency map of an image and the respective fixation map we calculated the symmetrized Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951) between the two (Rajashekar et al., 2004). The Kullback-Leibler divergence is an information theoretical measurement for the difference between two probability density functions (pdfs), in our case  $s(x, y)$  and  $f(x, y)$ :

$$D(s(x, y), f(x, y)) := D(s, f) = \sum_x \sum_y s(x, y) \log \frac{s(x, y)}{f(x, y)}$$

$D$  is always non-negative, and is zero, if and only if  $s(x, y) = f(x, y)$ . The smaller  $D$ , the higher the similarity between the two pdfs, with its lower bound at zero, if the two pdfs are identical. The so defined divergence happens to be asymmetric, that is  $D(s, f) \neq D(f, s)$ , for  $s \neq f$ . To circumvent an asymmetry of the measure for  $s \neq f$ , we chose the normalization method proposed by Johnson and Sinanovic, 2001:

$$KLD(s(x, y), f(x, y)) = KLD(s, f) = \frac{1}{\frac{1}{D(s, f)} + \frac{1}{D(f, s)}}$$

The smaller the KLD, the higher the similarity between the two pdfs, with its lower bound at zero, if the two pdfs are identical.

We defined  $KLD_{act}$  as the divergence between the saliency map and the fixations map. Under the experimental hypothesis this divergence should be small. To evaluate the significance of the measured, actual  $KLD_{act}$  we calculated the  $KLD$  -distributions under the assumption of independence of the two maps. One entry in this distribution was calculated as the distance  $KLD_{ind}$  between the original saliency pdf  $s(x, y)$  and a fixation map  $f(x, y)_{ind}$  derived from randomly (homogenously) distributed fixation points on the image (same number as were present in the original viewing, Parkhurst et al., 2002). This procedure was repeated 1000 times to yield the  $KLD_{ind}$  - distribution that served for testing whether the original viewing behavior measured by the actual  $KLD_{act}$  deviates

significantly from a viewing behavior that is not related to the saliency map (Figure 4B shows three examples).

For visualization purposes (Figure 4C) we show for each image the difference of the actual  $KLD_{act}$  value and the mean  $\langle KLD_{ind} \rangle$  of the  $KLD_{ind}$  -distribution:  $\Delta KLD = \langle KLD_{ind} \rangle - KLD_{act}$ . Positive values of  $\Delta KLD$  (i.e.,  $KLD_{act} < \langle KLD_{ind} \rangle$ ) denote a higher similarity between the actual fixation and saliency map than expected by a random viewer, indicating that the saliency map is a good predictor for the eye movements. On the contrary, negative values of  $\Delta KLD$  (i.e.,  $KLD_{act} > \langle KLD_{ind} \rangle$ ) signify that the distance between the actual fixation map and the saliency map is larger than assuming random viewing. Significant deviations from random viewing are marked by an asterisk.

In a further test, we repeated the whole above analysis considering fixations within ROIs only, and fed their number to the generator of random fixations (random viewer). The previous results were confirmed, i.e., significantly smaller  $KLD_{act}$  values for non-primate images, and significantly larger  $KLD_{act}$  values for primate images than expected (not shown).

#### 4.6 Spatial clustering of fixation positions

In order to investigate the existence of regions-of-interests (ROIs), defined as areas with high density of fixation positions, we identified spatial clusters of fixations by use of the mean shift algorithm (Fukunaga and Hosteler, 1975; Comaniciu and Meer, 2002) adapted for eye movement data (Santella and DeCarlo, 2004). This is an automatic, entirely data-driven method that derives the number and arrangement of clusters deterministically.

The algorithm starts from the set of  $N$  fixation positions  $\vec{v}_{i,j} = \begin{bmatrix} x_{i,j} \\ y_{i,j} \end{bmatrix}$ , with  $i \in (1, \dots, N)$  being the index of the fixation positions, and  $j = \mathbf{1}$  the original fixation positions on the 2D screen. The clustering algorithm proceeds iteratively, while moving at each iteration each of the points to its new

position  $\vec{V}_{i,j+1}$ , in dependence on the weighted mean of proximity and density of points around the

reference point,  $\vec{V}_{i,j+1} = \frac{\sum_i K(v_{i,j}-v_{k,j})v_{k,j}}{\sum_i (v_{i,j}-v_{k,j})}$  with  $j \neq k$ . The kernel  $K$  was defined as a 2D-Gaussian

with mean and variance of 0:  $K(\vec{v}) = \frac{e^{-(x^2+y^2)}}{\sigma^2}$ .  $\sigma$  was the only parameter of the clustering algorithm

and defined the attraction radius of the points. We varied its value and found 2.5 to yield satisfying

results, i.e., the algorithm did not lead to over fitting or to coarse clusters. We used this value to

perform all of our analyses. At each iteration the positions were moved into denser configurations, and

the procedure was stopped after convergence. Thereby fixations were assigned to a cluster whose

reference points lay within a diameter of  $1^\circ$  apart, referred to as experimental cluster. Robustness to

extreme outliers was achieved by limiting the support of points at large distances as defined by the

kernel  $K(\vec{v})$ .

In order to discard outlier clusters, we additionally applied a significance test to disregard clusters containing only a very small fraction of the data that deviate from expectation of independence.

As a significance test on the experimental clusters, we proceeded as follows: we assigned  $n$  random

locations on the screen by drawing  $n$  pairs of uniformly distributed numbers, with  $n$  being the total

number of fixations on a specific image. This random fixation map was fed into the mean shift

clustering algorithm, leading to a set of simulated clusters. Repeating this procedure 100 times, we

obtained two distributions: one of fixation numbers per cluster and one of cluster point density. An

experimental cluster of fixation positions was identified as significant when both its number of

fixations and their density exceeded the mean plus two standard deviations of the mean cluster point

number and density obtained from simulated clusters, respectively. All fixations that did not belong to a

significant cluster were pooled into a special cluster, referred to as background state. The background

state was crucial for the correct calculation of the probability of transitions to and from significant

clusters, i.e., in order to account also for the transitions that are neither within a cluster, nor between two clusters. Further details are described in the next section.

#### 4.7 Transition probabilities between fixation clusters

The statistical properties of the scanpaths a monkey chose to explore an image were analyzed by a Markov chain (MC) analysis (Markov, 1913). A MC is a sequence of random variables that propagate through a chain of states in accordance with given transition probabilities. These were estimated from the data as normalized frequencies of transitions from a specific state  $s_j$  to any particular other state  $s_k$  or to itself. The formerly identified clusters (compare previous section) of fixation points (including the background cluster) defined the states  $s_j$ . The transition probabilities from any one state to any other state (including the same state) were represented in matrix form. The state of the system at step  $t$  with  $t=1, \dots, T-1$ , with  $T$  being the total number of fixations on an image was derived via  $P(S_{t+1} = s | S_t = s_i, \dots, S_1 = s_1) = P(S_{t+1} = s | S_t = s_i)$  for all  $n$  states  $s_i \in s_1, \dots, s_n$ , thereby assuming that the scanpaths of the monkeys satisfy the Markov property, i.e., the present state is independent of the past states.

For better intuition, we visualized the results of the MC analysis by a transition graph (see example shown for monkey D in Figure 5), in which the vertices are the states, i.e., the identified fixation clusters. The graph is composed of oriented edges connecting vertices, weighted with the transition probabilities between the respective states. In addition, each vertex also contains an edge to itself weighted by the probability of staying within the same state in the subsequent step. In the following two cases no edges were drawn between the two vertices: first, whenever the transition probability  $P_{i,jk}$  equals zero; second, for transitions originating in the background state. For better visualization we represented the transition probabilities by the thickness of the edges (Figure 5C)

(thereby deviating in the graphical display from conventional transition graphs).

In order to interpret the transition probabilities derived by the MC analysis we compared them to the transition probabilities obtained assuming homogeneous chance probabilities of the transitions between any two states  $s_j$  and  $s_k$ ,  $P^{\text{expected}}(S_{t+1} = s_k | S_t = s_j) = P^{\text{expected}}(S_{t+1} = s_k) = \frac{n_k}{T}$ , with  $n_k$  being the number of fixations in state  $s_k$  and  $T$  the total number of transition steps. As illustrated in Figure 5C, we typically observed large differences in the transition probabilities in the viewing behavior of the monkeys (numbers in black) as compared to random transitions (numbers in gray). The monkeys' behavior displays a larger probability to stay within a significant cluster (Figure 5D), and a lower probability of moving to another significant cluster (Figure 5E) than a 'random viewer'. (Note, that the transition probabilities within and from the background cluster do not enter in the latter analysis.) This result holds true for both monkeys, and for images both containing and not containing faces.

In a second statistic we compared the transition probabilities obtained with the MC analysis with expected probabilities of staying within or switching between clusters weighted by the actual saccade length probabilities. This was obtained by multiplying the latter probabilities with the expected relative probability of transition (Figure 5D, E shown in green).

The expected transition probability between state  $s_j$  and  $s_k$  is:

$$P^{\text{expected}}(S_{t+1} = s_k | S_t = s_j) = \sum_d P^{\text{act}}(S_{t+1} = s_k | S_t = s_j; d) \cdot \rho_d^{j \rightarrow k},$$

with  $d$  being the saccade length and  $\rho_d^{j \rightarrow k}$  defined for cluster  $j$  as  $\rho_d^{j \rightarrow k} = \frac{P_d^{\text{theor}}(S_{t+1} = s_k | S_t = s_j; d)}{\sum_i P_d^{\text{theor}}(S_{t+1} = s_i | S_t = s_j; d)}$ ,

$\sum_k \rho_d^{j \rightarrow k} = 1, \forall (d, j)$ . The above probability that a saccade of length  $d$  leads to a state transition

$s_j \longrightarrow s_k$ ,  $P_d^{\text{theor}}(S_{t+1} = s_k | S_t = s_j; d)$ , was calculated from the obtained fixation clusters by

numerically computing all possible saccades of length  $d$  that stay within the same cluster  $s_j$  or land

into another cluster  $s_k$ .

Acknowledgements: We thank Tilke Judd (CSAIL MIT), Marc-Oliver Gewaltig and Ursula Körner (both HRI Europe), for stimulating discussions. Partially supported by the Stifterverband für die Deutsche Wissenschaft; Iniciativa Científica Milenio to PM and FJF; CONYCIT fellowship to FJF; the BMBF (grant 01GQ0413 to BCCN Berlin); and HRI, Europe.

## References

- Anderson, J. R., 1998. Social stimuli and social rewards in primate learning and cognition. *Behav. Process.* 42, 159–175.
- Andrew, R. J., 1963. Evolution of facial expression. *Science.* 142, 1034–1041.
- Antes, J. R., 1974. The time course of picture viewing. *J. Exp. Psychol.* 103, 62-70.
- Biedermann, I., 1987. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94, 115–147.
- Birmingham, E., Bischof, W.F. & Kingstone, A., 2009. Saliency does not account for fixations to eyes within social scenes. *Vis. Res.* 49, 2992–3000.
- Bruce, V., Young, A. 1998. In the eye of the beholder: the science of face perception. Oxford University Press, England, New York.
- Buswell, G., 1935. How people look at pictures: A study of the psychology of perception in art. Univ. Chicago Press, Chicago.
- Castelhana, M.S., Henderson M. S., 2007. Initial scene representations facilitate eye movement guidance in visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 753–763
- Cerf, M., Harel, J., Einhauser, W., Koch, C., 2008. Predicting human gaze using low-level saliency combined with face detection. *Advances in Neural Information Processing Systems*, 20, 241–248.
- Comaniciu, D., Meer P., 2002. Mean shift: A robust approach toward feature space analysis. *IEEE T. Pattern Anal.* 24, 603–619.
- Crick, F., Koch, C., 1998. Constraints on cortical and thalamic projections: The no-strong-loops hypothesis. *Nature.* 391, 245-250.
- Emery, N.J., 2000. The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604.
- Feng, G., 2006. Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research*, 7, 70–95.
- Funkunaga, K., Hosteler L., 1975. The estimation of the gradient of a density function, with

applications to pattern recognition, *IEEE T. Inform. Theory.* 21, 32–40.

Guo K, Robertson RG, Mahmoodi S, Tadmor Y, Young MP., 2003. How do monkeys view faces? –A study of eye movements. *Exp. Brain. Res.* 150, 363–374.

Guo K., Mahmoodi S., Robertson R.G., Young M.P., 2006. Longer fixation duration while viewing face images. *Exp. Brain. Res.* 171, 91-98.

Houtkamp R., Roelfsema P., 2010. Parallel and serial grouping of image elements in visual perception. *J. Exp. Psychol. Hum. Percept. Form.* 36, 1443-59.

Hubel, D.H., Wiesel, T. N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)*. 160, 106–154.

Itti, L., Koch, C., 1999. A comparison of feature combination strategies for saliency-based visual attention systems. *P. Soc. Photo-Opt. Ins.* 3644, 473–482.

Itti L., Koch, C., 2001. A model of saliency-based visual attention for rapid scene analysis. *Nature.* 2, 194–203.

Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE T. Pattern. Anal.* 20, 1254–1259.

Ito, J., Maldonado, P E., Singer, W., Grün, S., 2011. Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes. *Cereb. Cortex.* (in press).

Johnson, D., Sinanovic, S., 2001. Symmetrizing the Kullback Leibler distance. *IEEE Transactions on Information Theory.* 1, 1-10.

Judd, T., Durand, F., Torralba, A., 2009. Learning to Predict Where Humans Look. *IEEE 12<sup>th</sup> I. Conf. Comp. Vis.* pp. 2106-2113

Judge, S.J., Richmond, B.J., Chu, F.C., 1980. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res.* 20, 535–538.

Just, M. A., Carpenter, P. A., 1967. Eye fixations and cognitive processes. *Cognitive Psychol.* 8, 441–480.

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

Körner, E., Gewaltig, M.-O., Körner, U., Richter, A., Rodemann, T., 1999. A model of computation in neocortical architecture. *Neural Networks.* 12, 989–1005.

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.

Kullback, S., Leibler, R. A., 1951. On information and sufficiency. *Ann. Math. Stat.* 22, 79–86.

Mackworth, N. H., Morandi, A., 1967. The gaze selects informative details within pictures. *Percept. Psychophys.* 2, 547–552.

Maldonado, P., Babul, C., Singer, W., Rodriguez, E., Berger, D., Grün, S., 2008. Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images. *J. Neurophysiol.* 100, 1523–1532.

Markov, A., 1913. An example of statistical study on text of Eugeny Onegin illustrating the linking of events to a chain. *Izvestiya Akademii Nauk.* 6, 153–162.

- Mendelson, M., Haith, M., Goldman-Rakic, P., 1982. Face scanning and responsiveness to social cues in infant rhesus monkeys. *Dev. Psychol.* 18, 222–228.
- Milanse, R. Gil, S., Pun, T., 1995. Attentive mechanisms for dynamic and static scene analysis. *Opt. Eng.* 34, 2428–2434.
- Niebur, E., Koch, C., 1994. A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *J Comput Neurosci.* 1, 141–158.
- Noton, D., Stark, L., 1971a. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Res.* 11, 929–942.
- Noton, D., Stark, L., 1971b. Scanpaths in eye movements during pattern perception. *Science.* 171, 308–311.
- Pannasch, S., Helmert, J. R., Roth, K., Herbold, A.-K., Walter, H., 2008. Visual fixation durations and saccadic amplitudes: Shifting relationship in a variety of conditions. *J. Eye Mov. Res.* 2:4, 1-19.
- Parkhurst, D., Law K., Niebur E., 2002. Modeling the role of saliency in the allocation of overt visual attention. *Vision Res.* 42, 107-123.
- Rajashekar, U., Cormack, L. K., Bovik, A. C., 2004. Point of gaze analysis reveals visual search strategies. In Rogowitz B. E., Pappas T. N. (Eds.), *Proceedings of SPIE Human Vision and Electronic Imaging IX* (vol. 5292, pp. 296–306). Bellingham, WA: SPIE Press.
- Rosenfeld, S. A., Van Hoesen, G. W., 1979. Face recognition in the rhesus monkey. *Neuropsychologia.* 17, 503–509.
- Santella, A., DeCarlo, D., 2004. Robust clustering of eye movement recordings for quantification of visual interest. In: *ETRA Proceedings 2004*. ACM, New York, pp. 27–34.
- Shepherd, S. V., Steckenfinger, S. A., Hasson, U., Ghazanfar, A. A., 2010. Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Curr. Biol.* 20, 649-656.
- Simola, J., Salojärvi, J., Kojo, I., 2008. Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research.* 9, 237–251.
- Stratton G M, 1902 “Eye-movements and the aesthetics of visual form” *Philosophische Studien* 20, 336–359.
- Stratton G M, 1906 “Symmetry, linear illusions and the movements of the eye” *Psychological Review* 13 82–96.
- Torralba, A, Oliva A, Castelhana M.S., Henderson J.M., 2006. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 113, 766-786.
- Thorpe, S., Fize, D., Marlot., C., 1996. Speed of processing in the human visual system. *Nature.* 381, 520-522.
- Treisman, A.M., Gelade G., 1980. A feature-integration theory of attention. *Cogn. Psychol.* 12, 97-136.
- Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y. H., Davis, N., Nuflo, F., 1995. Modeling visual attention via selective tuning. *Artiff. Intell.* 78, 507–545.
- Ullman, S., 1995. Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb. Cortex.* 5, 1–11.
- Unema, P. J. A., Pannasch, S., Joos, M., Velichkovsky, B. M., 2005. Time course of information

- processing during scene perception: The relationship between saccade amplitude and fixation duration. *Vis. Cogn.* 12, 473-494.
- Van Der Lans, R., Pieters, R., Wedel, M., 2008. Eye-Movement Analysis of Search Effectiveness. *Journal of the American Statistical Association.* 103, 452-461.
- Van Rullen, R., Gautrais, J., Delorme, A., Thorpe, S., 1998. Face processing using one spike per neurone. *Biosystems.* 48, 229–239.
- Walther, D. Koch, C., 2006. Modeling attention to salient proto-objects. *Neural Networks.* 19, 1395–1407.
- Wedel, M. Pieters, R., Liechty, J. 2008. Attention switching during scene perception: how goals influence the time course of eye movements across advertisements. *Journal of Experimental Psychology-Applied*, 14, 129-138.
- Wolfe, J.M., 1994. Guided search 2.0: A revised model of visual search. *Psychonomic Bull. Rev.* 1, 202–238.
- Yarbus, A.L., 1967. *Eye movements and vision.* Plenum Press. New York.
- Yen, S.C., Baker, J., Lachaux, J.P., Gray, C.M., 2008. Responses of primary visual cortical neurons to natural movies in anesthetized cat. *BMC Neurosci.* 9, P125.

Figure 1: Time course of the experiment. Images (IMAGE), blank screen (BLANK) and blank screens with a fixation spot (FIXATION CUE) were presented in an interleaved manner, with pause of 2-3 sec in between. For each of these presentations, the monkeys had to keep their gaze within the limits of the screen for 3-5 sec. In case the presentation of FIXATION CUE they had to fixate the cue for 1 sec. Successful behavior was rewarded with a drop of juice. Within one experimental session a set of 4-7 different images (out of 11, randomly selected) were presented in random manner. The presentations covered the size of the screen ( $30 \times 40$  cm), which corresponded to a visual angle of  $30 \times 40$  degrees since presented 57 cm in front of the monkeys.

Figure 2: Statistics of the durations of fixations and saccades. A. Example of one presented image (image no. 3) with the scanpath (white lines) during one single image presentation. The red dots indicate the positions of fixations, the numbers indicate their sequences. B. Time course of the eye positions during the image presentation (upper trace: horizontal, lower trace: vertical). Fixations are marked in red, saccades in green, and unclassifiable events are colored in black. C. Distributions of fixation durations (bin width: 25 msec) and D. saccade durations (bin width: 5 msec), pooled over all sessions and all image presentation, shown separately for each of the three monkeys.

Figure 3: Spatial distribution of fixations on four example images. Each row corresponds to one of four example images shown interleaved with pure fixation distributions. For each monkey we show two columns, the left illustrates the pure fixation positions, and in the right the fixation positions are overlaid onto the presented image. The displayed fixation positions in each subfigure were collected from all presentations of the respective image, over all experimental sessions. Bottom row: Merged fixation positions for each monkey on all presented images from all sessions.

Figure 4: Correlations between saliency and fixation maps. A. Saliency maps of three example images shown in combination with fixation positions of monkey D (green dots) on these pooled from all sessions. B. KLDs resulting from correlating the saliency maps and fixation maps of the same data as in A.). The KLDs derived from the original data are marked by vertical lines, the KLD distributions of random fixation positions (1000 repetitions) are shown as histograms. The different colors mark the three example images shown in A.. C. Differences ( $\Delta$  KLD) of actual KLD and mean of KLDs of random fixation positions shown for the different images (below). Colors of the bars refer to the  $\Delta$  KLDs of the different monkeys. Asterisks mark  $\Delta$  KLDs that deviate significantly from random viewers ( $P < 0.01$ ). Green X mark images that had not been presented to monkey M.

Figure 5: Transition probabilities between clusters of fixations of monkey D viewing image no. 2. A. Example of eye movement trajectories during 4 different presentations of the same image (different colors). B. Spatial clusters (3) of fixation positions extracted by the mean shift algorithm. The fixation points in the respective clusters are colored in red, cyan, and magenta. C. Transition graphs between the identified clusters, now interpreted as the states of the MC model (same colors as in B.). Fixations that do not fall into any of the identified clusters are collected into the background state (black box). Experimental transition probabilities are indicated in black, transition probabilities assuming a random viewer are indicated in gray. STOP refers to eye movements exceeding the image presentation time. D. Distributions of probabilities of staying within the same cluster for experimental (black), random viewer (red), and predicted on the basis of the saccade amplitude distribution (green), for the same data as shown in A.-C.. E. Distributions of probabilities to switch between clusters for experimental (black), random viewer (red), and predicted on the basis of the saccade amplitude distribution (green), for the same data as shown in A.-C..

Fig 1

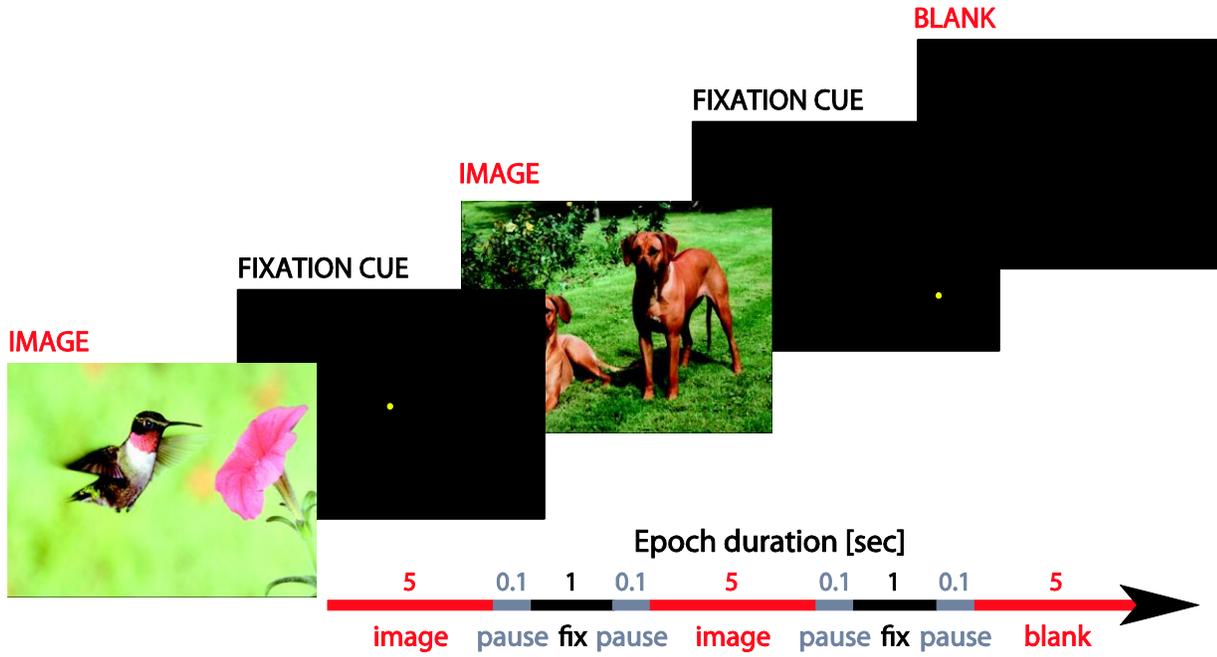


Fig 2

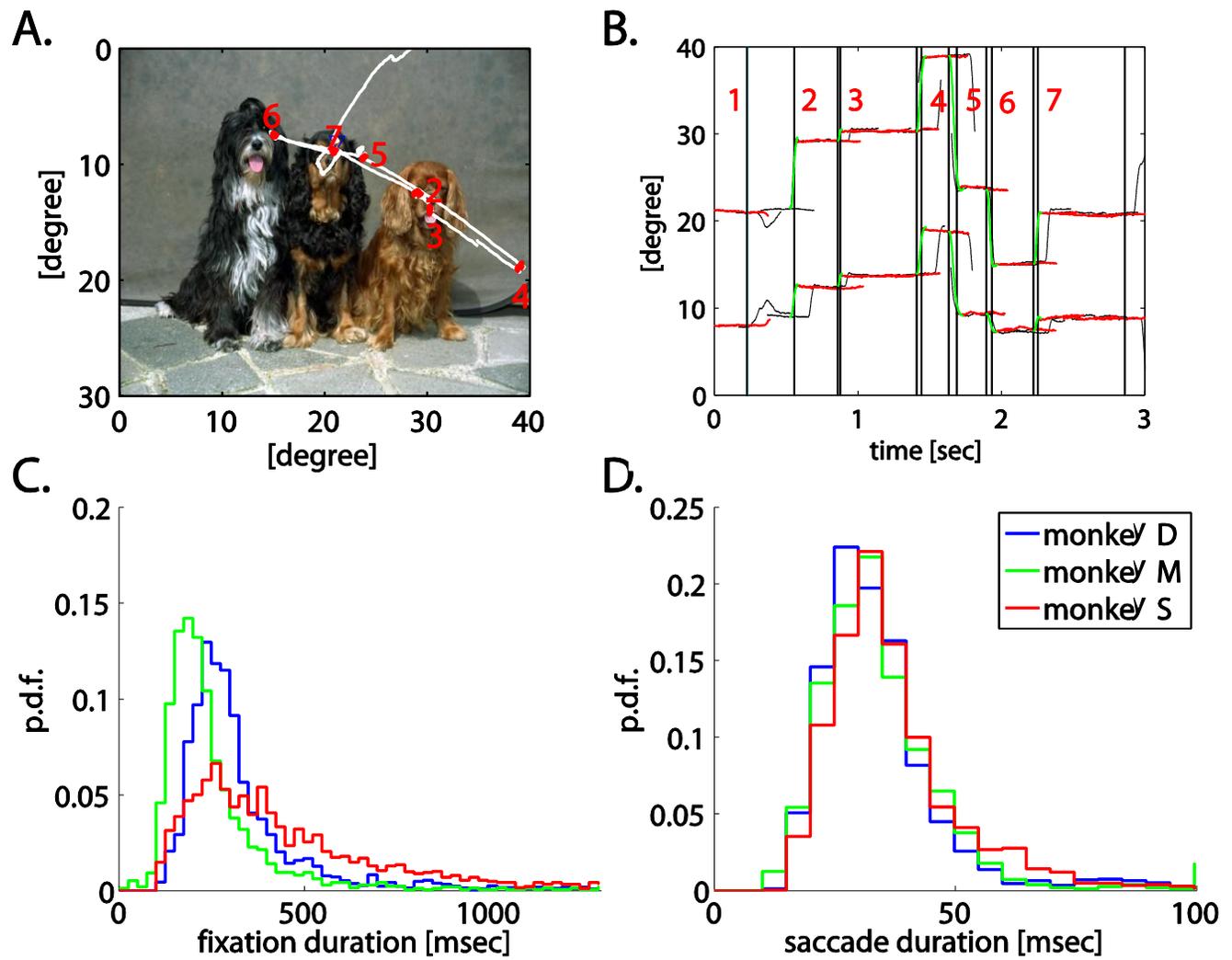


Fig 3

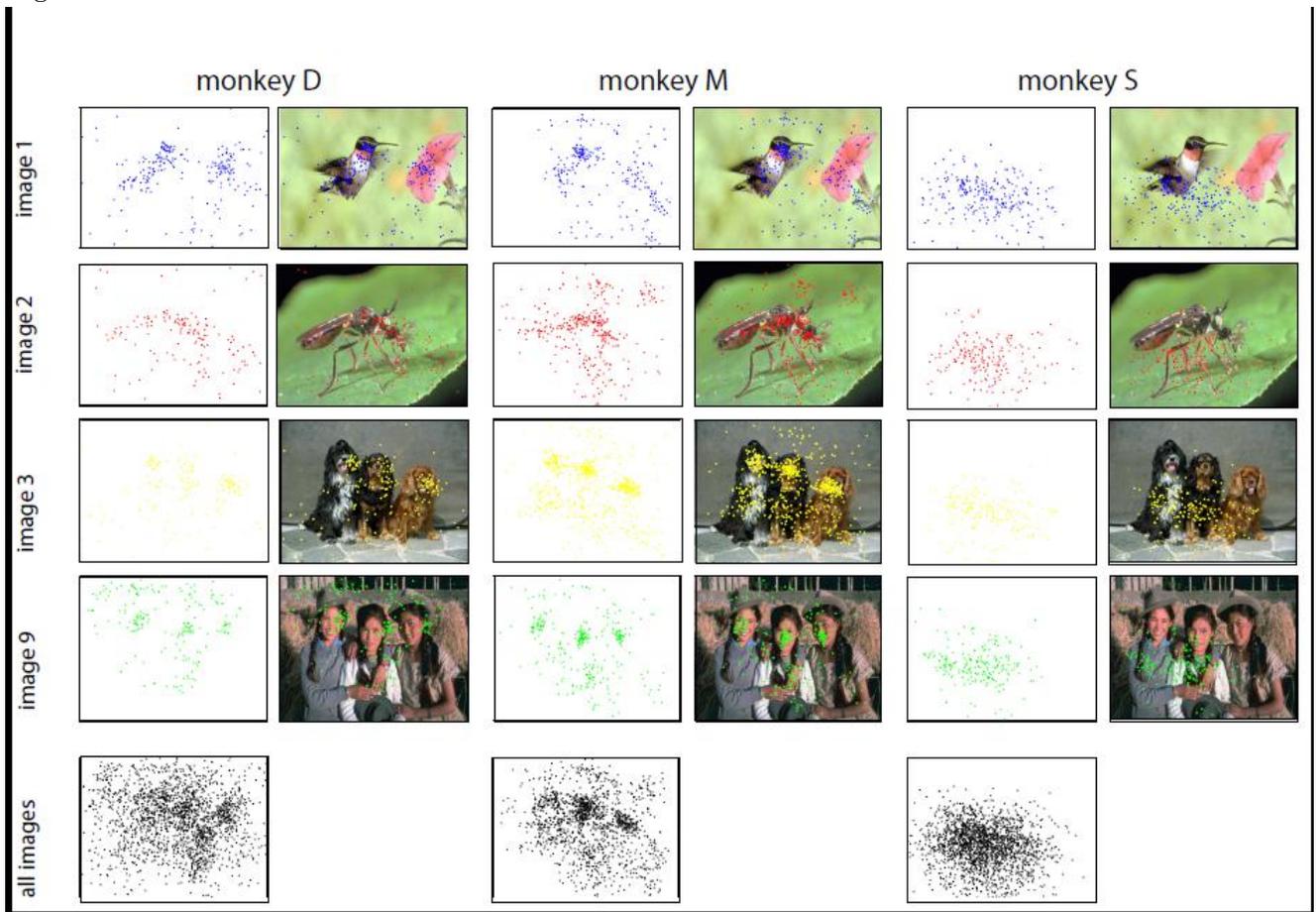


Fig 4

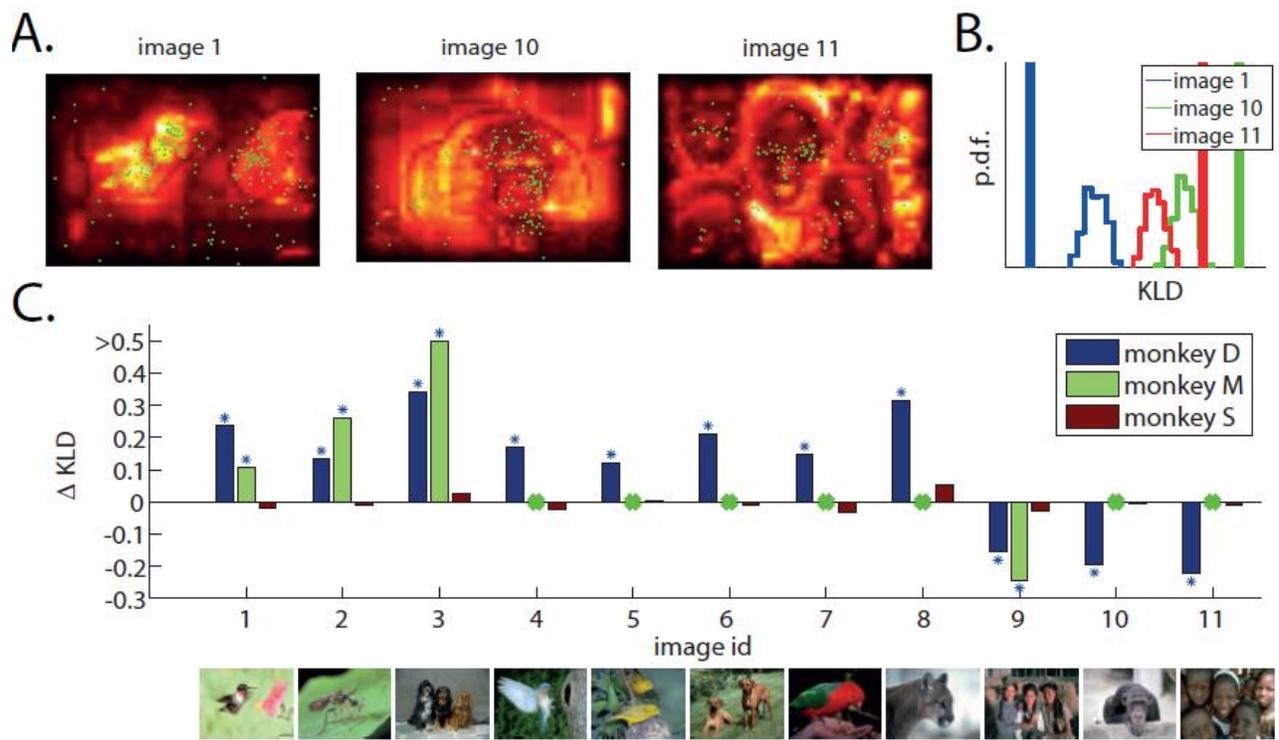


Fig 5

Figure

