BRAIN RESEARCH 1434 (2012) 34–46

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

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

*Denise Berger*[a,b,*,1], *Antonio Pazienti*[c], *Francisco J. Flores*[d,2], *Martin P. Nawrot*[a,b], *Pedro E. Maldonado*[d], *Sonja Grün*[e,f]

[a]*Neuroinformatics, Institute of Biology, Freie Universität, Königin-Luise Str. 1-3, Berlin, Germany*
[b]*Bernstein Center for Computational Neuroscience, Phillipstr. 13, Berlin, Germany*
[c]*European Brain Research Institute, Via del fosso di Fiorano 64, Rome, Italy*
[d]*Program of Physiology and Biophysics, Faculty of Medicine, Universidad de Chile, Av. Independencia 1027, Santiago, Chile*
[e]*Institute of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Research Center Jülich, 52425 Jülich, Germany*
[f]*Theoretical Systems Neurobiology, RWTH Aachen University, 52056 Aachen, Germany*

A R T I C L E  I N F O

*Article history:*
Accepted 7 October 2011
Available online 14 October 2011

*Keywords:*
Eye movement
Scan path
Fixation map
Natural vision
Monkey

A B S T R A C T

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, independent 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.



* *Corresponding author at:* Laboratory of Neuromotor Physiology, IRCCS Fondazione Santa Lucia, Via Ardeatina 306, 00179 Roma, Italy. Fax: +39 0651501482.

E-mail addresses: d_berger@gmx.net (D. Berger), apazienti@gmx.net (A. Pazienti), fjflores@mit.edu (F.J. Flores), martin.nawrot@fu-berlin.de (M.P. Nawrot), pedro@neuro.med.uchile.cl (P.E. Maldonado), s.gruen@fz-juelich.de (S. Grün).

[1] Present address: Laboratory of Neuromotor Physiology, IRCCS Fondazione Santa Lucia, Via Ardeatina 306, 00179 Roma, Italy.

[2] 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.

0006-8993 
doi:10.1016/j.brainres.2011.10.013

## 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 (Biedermann, 1987; Crick and Koch, 1998; Noton and Stark, 1971a), 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 (Itti and Koch, 1999, 2001; Koch and Ullman, 1985; Treisman and Gelade, 1980) while top–down factors depend on the task being executed during exploration of a scene (Buswell, 1935; Just and Carpenter, 1967; Yarbus, 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., Itti et al., 1998; Milanse et al., 1995; Tsotsos et al., 1995; Wolfe, 1994). The most common strategy includes the computation of saliency maps to account for bottom–up factors and defines 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 (Itti et al., 1998; Milanse et al., 1995). 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

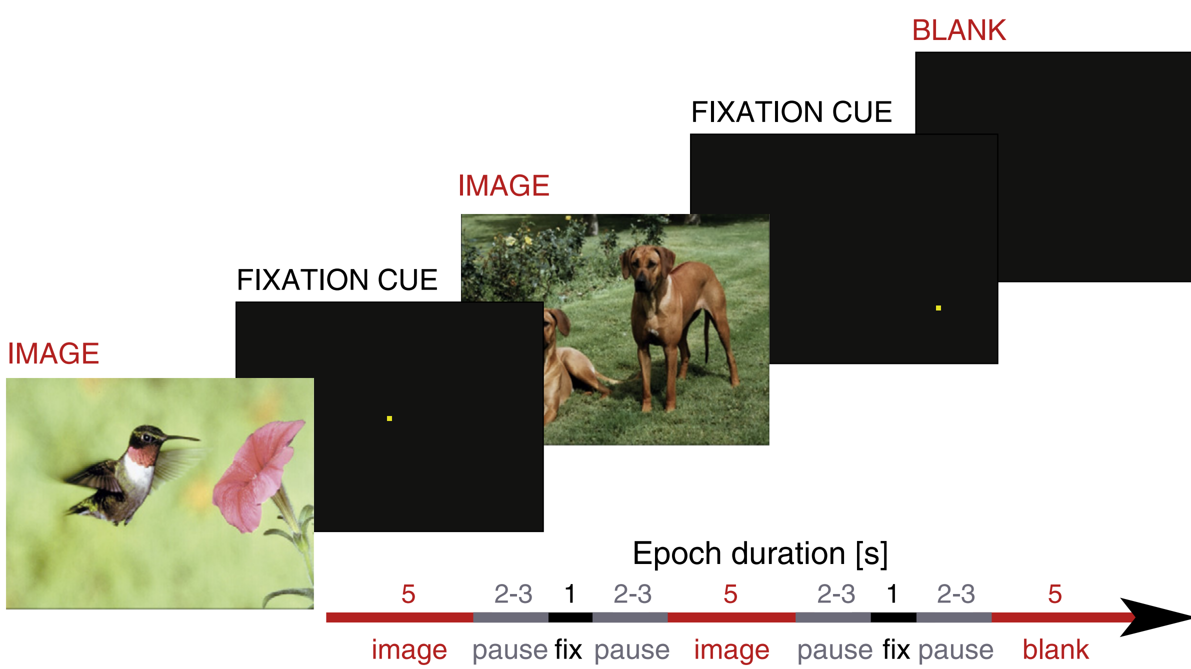


**Fig. 1 – Time course of the experiment. Images (IMAGE), blank screen (BLANK) and blank screens with a fixation spot (FIXATION CUE) were presented in an interleaved manner, with a pause of 2–3 s in between. For each of these presentations, the monkeys had to keep their gaze within the limits of the screen for 3–5 s. In case the presentation of FIXATION CUE they had to fixate the cue for 1 s. 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 × 40 cm), which corresponded to a visual angle of 30 × 40° since presented 57 cm in front of the monkeys.**