Available online at www.sciencedirect.com
ScienceDirect
Neural Networks 19 (2006) 1383–1394
Neural Networks
www.elsevier.com/locate/neunet

2006 Special Issue

Attention in natural scenes: Neurophysiological and computational bases

Edmund T. Rolls[a,*], Gustavo Deco[b]

[a] University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, United Kingdom
[b] Institució Catalana de Recerca i Estudis Avançats (ICREA), Universitat Pompeu Fabra, Department of Technology, Computational Neuroscience, Passeig de Circumval.lació, 8, 08003 Barcelona, Spain

Received 11 May 2006; accepted 1 August 2006

Abstract

How does attention operate in natural scenes? We show that the receptive fields of inferior temporal cortex neurons that implement object representations become small and located at the fovea in complex natural scenes. This facilitates the readout of information about an object that may be reward or punishment associated, and may be the target for action. Top-down biased competition to implement attention has a much weaker effect in complex natural scenes than in otherwise blank scenes with two objects. Part of the solution to the binding problem is thus that competition and the foveal cortical magnification factor emphasize what is present at the fovea, and limit the binding problem. Part of the solution to the binding problem is that neurons respond to combinations of features present in the correct relative spatial positions. Stimulus-dependent neuronal synchrony does not appear to be quantitatively important in feature binding, and in attention, in natural visual scenes, at least in the inferior temporal visual cortex, as shown by information theoretic analyses. The perception of multiple objects in a scene is facilitated by the fact that inferior temporal visual cortex neurons have asymmetrical receptive fields with respect to the fovea in complex scenes. Computational models of this processing are described.

Keywords: Attention; Biased competition; Natural scenes; Receptive field size; Inferior temporal visual cortex

We have produced a unified theory of visual attention and working memory, and how these processes are influenced by rewards to influence decision making (Deco & Rolls, 2005a, 2006; Rolls & Deco, 2002). The theory is based on the *'Biased Competition Hypothesis'* (Chelazzi, 1998; Chelazzi, Miller, Duncan, & Desimone, 1993; Miller, Gochin, & Gross, 1993; Moran & Desimone, 1985; Motter, 1993; Reynolds & Desimone, 1999; Rolls & Deco, 2002; Spitzer, Desimone, & Moran, 1988). In this approach (Deco & Rolls, 2005a; Rolls & Deco, 2002) multiple activated populations of neurons which may be hierarchically organized engage in competitive interactions, and external top-down effects bias this competition in favor of specific neurons (see Section 4). This approach leaves open however how object and spatial attention operate in complex natural scenes, the neural encoding of information about objects, and how multiple objects and feature binding are implemented in complex scenes. These are the main subjects of this paper.

* Corresponding author. Tel.: +44 1865 271348; fax: +44 1865 310447.
E-mail address: Edmund.Rolls@psy.ox.ac.uk (E.T. Rolls).
URL: http://www.cns.ox.ac.uk (E.T. Rolls).

0893-6080/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2006.08.007

1. Object-based attention and object selection in complex natural scenes

1.1. Neurophysiology of object selection in the inferior temporal visual cortex

Object-based attention refers to attention to an object. For example, in a visual search task the object might be specified as what should be searched for, and its location must be found. In spatial attention, a particular location in a scene is pre-cued, and the object at that location may need to be identified.

Much of the neurophysiology, psychophysics, and modelling of attention has been with a small number, typically two, of objects in an otherwise blank scene. In this section, we consider how attention operates in complex natural scenes, and in particular describe how the inferior temporal visual cortex operates to enable the selection of an object in a complex natural scene. The inferior temporal visual cortex contains distributed and invariant representations of objects and faces (Booth & Rolls, 1998; Hasselmo, Rolls, & Baylis, 1989; Rolls, 2000, in press; Rolls & Baylis, 1986; Rolls & Deco, 2002; Rolls & Tovee, 1995; Rolls, Treves, & Tovee, 1997; Tovee, Rolls, & Azzopardi, 1994).

To investigate how attention operates in complex natural scenes, and how information is passed from the inferior temporal cortex (IT) to other brain regions to enable stimuli to be selected from natural scenes for action, Rolls, Aggelopoulos, and Zheng (2003) analyzed the responses of inferior temporal cortex neurons to stimuli presented in complex natural backgrounds. The monkey had to search for two objects on a screen, and a touch of one object was rewarded with juice, and of another object was punished with saline (see Fig. 1 for a schematic illustration and Fig. 2 for a version of the display with examples of the stimuli shown to scale). Neuronal responses to the effective stimuli for the neurons were compared when the objects were presented in the natural scene or on a plain background. It was found that the overall response of the neuron to objects was hardly reduced when they were presented in natural scenes, and the selectivity of the neurons remained. However, the main finding was that the magnitudes of the responses of the neurons typically became much less in the real scene the further the monkey fixated in the scene away from the object (see Fig. 3).

It is proposed that this reduced translation invariance in natural scenes helps an unambiguous representation of an object which may be the target for action to be passed to the brain regions which receive from the primate inferior temporal visual cortex. It helps with the binding problem, by reducing in natural scenes the effective receptive field of at least some inferior temporal cortex neurons to approximately the size of an object in the scene.

It is also found that in natural scenes, the effect of object-based attention on the response properties of inferior temporal cortex neurons is relatively small, as illustrated in Fig. 4 (Rolls, Aggelopoulos et al., 2003).

The results summarized in Fig. 4 for 5° stimuli show that the receptive fields were large (77.6°) with a single stimulus in a blank background (top left), and were greatly reduced in size (to 22.0°) when presented in a complex natural scene (top right). The results also show that there was little difference in receptive field size or firing rate in the complex background when the effective stimulus was selected for action (bottom right, 19.2°), and when it was not (middle right, 15.6°) (Rolls, Aggelopoulos et al., 2003). (For comparison, the effects of attention against a blank background were much larger, with the receptive field increasing from 17.2° to 47.0° as a result of object-based attention, as shown in Fig. 4, left middle and bottom.) The computational basis for these relatively minor effects of object-based attention when objects are viewed in natural scenes is considered in Section 1.2.

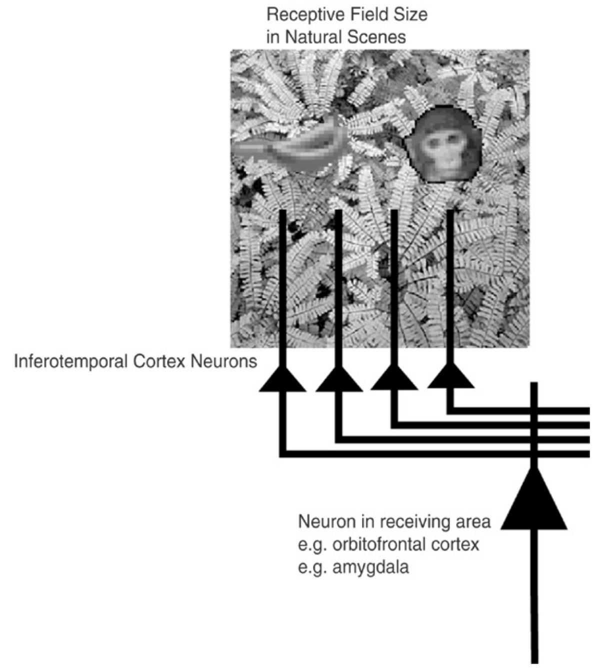

Fig. 1. Objects shown in a natural scene, in which the task was to search for and touch one of the stimuli. The objects in the task as run were smaller. The diagram shows that if the receptive fields of inferior temporal cortex neurons are large in natural scenes with multiple objects, then any receiving neuron in structures such as the orbitofrontal cortex and amygdala would receive information from many stimuli in the field of view, and would not be able to provide evidence about each of the stimuli separately.

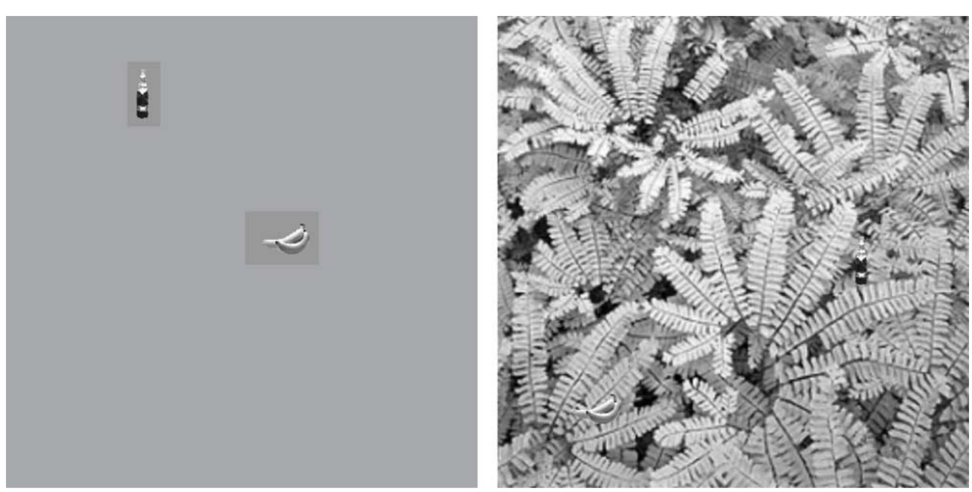

Fig. 2. The visual search task. The monkey had to search for and touch an object (in this case a banana) when shown in a complex natural scene, or when shown on a plain background. In each case a second object is present (a bottle) which the monkey must not touch. The stimuli are shown to scale. The screen subtended 70° × 55°.