Vision Research 104 (2014) 36–46

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Contextual modulation and stimulus selectivity in extrastriate cortex

Matthew R. Krause *, Christopher C. Pack

Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC, Canada

ARTICLE INFO

Article history:
Received 25 March 2014
Received in revised form 8 October 2014
Available online 3 November 2014

Keywords:
Contextual modulation
Normalization
Surround
Extrastriate cortex
Neurophysiology
Macaque

ABSTRACT

Contextual modulation is observed throughout the visual system, using techniques ranging from single-neuron recordings to behavioral experiments. Its role in generating feature selectivity within the retina and primary visual cortex has been extensively described in the literature. Here, we describe how similar computations can also elaborate feature selectivity in the extrastriate areas of both the dorsal and ventral streams of the primate visual system. We discuss recent work that makes use of normalization models to test specific roles for contextual modulation in visual cortex function. We suggest that contextual modulation renders neuronal populations more selective for naturalistic stimuli. Specifically, we discuss contextual modulation's role in processing optic flow in areas MT and MST and for representing naturally occurring curvature and contours in areas V4 and IT. We also describe how the circuitry that supports contextual modulation is robust to variations in overall input levels. Finally, we describe how this theory relates to other hypothesized roles for contextual modulation.

1. Introduction

Visual information is rarely found in isolation. A typical scene contains many objects, each of which can be defined by its own combination of visual features. Many of these features, such as orientation and spatial frequency, are extracted by dedicated mechanisms in the early visual system. These circuits are thought to generate feature selectivity in part by repeatedly filtering and pooling feedforward inputs. For example, a V1 neuron could develop orientation tuning by selectively pooling the outputs of several circular LGN receptive fields (Hubel & Wiesel, 1962); the LGN receptive fields, in turn, arise from filtering and pooling in the retina and retinal ganglion cells. Since the neurons implementing these operations have small spatial receptive fields and short memories, it may seem like processing should be quite local in space and time.

However, it has long been known that the processing of a visual stimulus is affected by the overall *gestalt*, or context, in which it occurs. The presence of a stimulus, even one that cannot directly drive a neuron's feedforward inputs (e.g., because it is outside the cell's spatial receptive field or its tuning passband), can affect how the cell responds to other stimuli that do engage its feedforward inputs.

A whimsical example of this effect can be found in Quiroga et al. (2005), who recorded the activity of medial temporal lobe (MTL) neurons in human patients while the patients viewed photographs. One neuron, shown in Fig. 1, responded strongly and almost exclusively to photos of the actress Jennifer Aniston (shaded regions in the top row), regardless of the low-level features (e.g., color, edge orientation) that comprise her portrait. This selectivity and invariance is common in high-level cortical areas (Desimone et al., 1984; Tsunoda et al., 2001), but virtually unheard of in lower ones, where cells respond to any stimulus containing an appropriate angle or hue (Hubel & Wiesel, 1968; Leventhal et al., 1995). However, there is virtually no response to images that contain both Aniston and her then-husband, actor Brad Pitt (Fig. 1, top-right), again regardless of the low-level features that make up his appearance. Thus, one might conclude that Brad Pitt's presence suppresses the cells' responses to Jennifer Aniston.

However, other models might also explain these responses. The response pattern might reflect selectivity for a specific, low-level feature (e.g., orientation or color) that happens to be present in all of the Aniston images but none of the Aniston + Pitt images. Or perhaps any stimulus accompanying Aniston, other than the background, leads to suppression. Because the stimulus features that activate MTL neurons are not well understood, there is little basis for estimating the contribution of these different mechanisms.

The difficulty in modeling such complex visual selectivity thus arises from the variety of possible inputs and, in many cases, from a lack of detailed knowledge of the computations performed by cortical neurons selective for complex stimuli. Recent work in this area has attempted to solve this problem by leveraging the

* Corresponding author at: Room 888, 3801 rue University, Montreal, QC H3A 2B4, Canada.
E-mail address: matthew.krause@mcgill.ca (M.R. Krause).

http://dx.doi.org/10.1016/j.visres.2014.10.006
0042-6989/

available knowledge on low-level visual processing. We suggest that a particularly fruitful approach is to present neurons with a large variety of stimuli that explore, to the extent possible, a wide range of feature combinations. The resulting data are then fit to models that approximate the hierarchical structure of the visual system (Brincat & Connor, 2004; Mineault et al., 2012; Rust et al., 2006). Thus, for example, models of V2 can be framed as performing computations on the outputs of simulated V1 neurons, rather than operating on the raw visual input (Coen-Cagli & Schwartz, 2013). The precise operations that are used typically include feedforward filtering, as well as contextual modulations such as normalization.

2. Contextual modulation in striate cortex

Contextual modulation is typically measured in neurophysiological experiments using a simple paradigm. Investigators isolate a neuron and map its classical receptive field (CRF). They then compare responses to stimuli placed only within the cell's CRF with those that extend beyond its boundaries. When the contents of the CRF are identical in the two conditions, any observed difference is then ascribed to contextual modulation, and the spatial area producing these effects is called the non-classical receptive field (nCRF), or surround.

Although there is some evidence for excitatory contextual modulation (Angelucci & Bressloff, 2006; Bringuier et al., 1999), the net effect of nCRF stimulation is typically suppressive: stimulating the nCRF with large, high contrast stimuli reduces V1 neurons' firing rates by 40–70%, compared to CRF-only stimulation (reviewed in Series et al. (2002)); similar results have also been obtained in extrastriate areas. Several functional roles have been proposed for this modulation, including the following:

- figure-ground segmentation (Allman, Miezin, & McGuinness, 1985),
- redundancy reduction (Atick & Redlich, 1990; Dong & Atick, 1995),
- generation of a sparse code (Vinje & Gallant, 2000),
- firing rate control/metabolic efficiency (Attwell & Laughlin, 2001), and
- noise rejection (Chen, Geisler, & Seidemann, 2006).

These hypotheses all share a common feature: contextual modulation is used to refine existing feature representations that have been generated by other—unspecified, but presumably feedforward—circuitry. Here we review evidence suggesting that contextual modulation can do more, and actually creates neural selectivity for new and complex visual features. There is almost universal agreement that this occurs in the retina, where contextual modulation—implemented through lateral inhibition—converts the absolute luminance information captured by the retina into a new image feature, local contrast (Hartline, 1940; Kuffler, 1953). This review focuses on the consequences of iterating similar mechanisms across multiple visual cortical areas, a topic that has been explored less thoroughly (but see Gautama & Van Hulle, 2001).

Contextual modulation is typically thought to arise from interactions between neurons. These interactions can take several forms. When expressed mathematically as a subtraction of two quantities, akin to the integration of IPSPs and EPSPs, the modulation is usually called opponent inhibition (Hurvich & Jameson, 1957; Reid & Shapley, 1992). Interactions between neighboring bipolar or amacrine cells, for example, are often described using opponent models. When these interactions are expressed using a divisive interaction between neurons—or populations of neurons—the resulting model is usually called a normalization model. These models have a long history in visual neuroscience

Fig. 1. Do MTL neurons exhibit complex contextual modulation? Quiroga et al. (2005) recorded the activity of a medial temporal lobe (MTL) neuron while human patients viewed images. This neuron responded vigorously (individual trials shown in center; peristimulus histograms shown in the bottom row) whenever the patient saw Jennifer Aniston, but was suppressed whenever Brad Pitt was also in the photograph. However, we know very little about MTL neurons' feature selectivity or the computations they perform that might evoke this suppression. Reprinted by permission from Macmillan Publishers Ltd.: Nature (Quiroga et al., 2005) © 2005.