Pre-Attentive Visual Selection

Li Zhaoping[a,*]   Peter Dayan[b]
[a]University College London, Dept. of Psychology, UK
[b]University College London, Gatsby Computational Neuroscience Unit, UK
[*]Correspondence to L. Zhaoping at z.li@ucl.ac.uk

In a special issue of *Neural Networks* devoted to attention, it might seem at the very least perverse to write about pre-attentive processing. However, it is only by understanding the power of pre-attentive processes that we can understand what attention has to work with, and indeed against.

In this note, we consider the critical function of the selection from an entire visual scene of visual locations or objects for detailed or attentive processing. Such selection of some places at the expense of others is necessary because attention has only a meagre capacity (estimated at just 40 bits/second; Sziklai 1956). The same bottleneck implies that the selection process itself cannot generally be attentive (bar an explicit cue or some other form of effective guidance). Fortunately, pre-attentive mechanisms, operating in parallel across the entire visual input, seen most prominently in pop-out (with, for instance, a red dot popping out among the green ones, or a vertical bar among horizontal bars), operate to offer a vastly simpler substrate of salience for attentional selection (Neisser 1967; Triesman & Gelade, 1980, Julesz 1981). Concomitantly, tasks that require the selection of non-cued locations that are not pre-attentively salient, are much more daunting.

What clues are there to the mechanisms employed by pre-attentive selection? The computational requirements indicate (a) fast parallel processing, (b) a spatial substrate for selection, and (c) a complex featural basis of effects such as pop-out. These particularly implicate V1, as the largest, and retinotopic, cortical visual area, with many cells tuned to different visual feature(s) responding at each location in visual space. Physiological data (Allman et al, 1985, Knierim and Van Essen 1992, Sillito et al 1995, Nothdurft et al 1999) suggest that V1 contributes to pop-out, since a V1 neuron's response to a pop out item is higher than to a background item, for both anesthetized and awake animals.

On the basis of these and other findings, it has recently been proposed (Li 1999a;b, 2002, Zhaoping 2005) that V1 creates a general, pre-attentive saliency map, with the receptive field location of the most active V1 neuron responding to a scene most likely to be selected, regardless of the feature preferences of the V1 cells concerned. This theory suggests that the computation of salience is instantiated in the neural dynamics arising from the horizontal, intra-cortical interactions between neurons that typically arrange for the suppression of a V1 neuron's response by activities of nearby neurons tuned to similar features (Knierim and van Essen 1992, Wachtler et al 2003), i.e., iso-feature suppression (which is effective within 10-20ms after the initial responses). Accordingly, neural activities come to highlight the breakdown of statistical homogeneity in the input, occurring typically at locations of pop out items or visual surface borders.

This saliency map can then be read-out in a very quick and feature-blind manner, by a down stream area such as the superior colliculus, which is involved in eye movements (Tehovnik, Slocum & Schiller, 2003), using the firing rates of the neurons as a universal currency for bottom-up saliency in a process of bidding for visual selection. Consequently, the chance of a location being selected is determined by the maximum of the activities of all neurons sharing that location as their receptive field. Since the lateral neural connections, mediating horizontal intra-cortical interactions, depend sensitively on pre- and post-synaptic feature selectivities (Gilbert and Wiesel 1983, Rockland and Lund 1983), potentially in a plastic manner, the computation of pre-attentive preference can be highly sophisticated with respect both to the statistics of a single image, and the whole collection of images in a subject's overall visual diet. Further, since many V1 neurons are tuned to more than one feature dimension, the processing of salience is coupled between different dimensions.

According to this proposal, the activities of the V1 neurons determining the salience of each location depend on the input at those locations as well as the input in the spatial context. This means that the ultimate salience at an input location (remembering that only the maximal activity matters) can depend on different aspects of the input under different circumstances, as should be the case for saliency. To be concrete, the salience of a red-vertical bar may be signalled by a red-tuned cell if this bar is among green-vertical bars, but by a different neuron with the same spatial receptive field, but tuned to vertical orientation, if the bar is in a context of red horizontal bars. To a red-vertical bar, the response of the vertical tuned cell is suppressed in a green-vertical context due to iso-orientation suppression, and the response of the red tuned cell is suppressed in a red-horizontal context due to iso-color suppression, making the red-tuned or vertical-tuned cell, respectively be the most active, and thus signal saliency.

The most parsimonious version of the theory suggests that bottom-up selection is totally blind to the less salient feature at each location (in our example, the vertical or red feature respectively), that do not elicit the highest response at this location. If we assume that pressure on subjects to respond quickly in visual tasks forces them to place greater reliance on pre-attentive mechanisms, this prediction can be tested by examining the consequences of making either the more or less salient features relevant or irrelevant to a task.

Zhaoping and May (2004) conducted an experiment (illustrated in Fig. 1) exactly along these lines. They showed that segmenting two homogeneous textures of left tilt and right tilt bars is made much harder if task-irrelevant (distractor) horizontal and vertical bars in a checker-board pattern are superposed. Without the distractors, the bars at the texture border elicit higher responses than those in the background, since they have fewer iso-orientation neighbors. This makes the texture border pre-attentively salient, rendering a fast decision easy. However, the distractor bars also lack iso-orientation neighbors, and so are just as salient as the bars in the texture border. Thus, when distractors are present, all locations are equally pre-attentively salient, making the border very hard to find. That the border is hard to find goes against models encompassing pre-attentive salience that sum up the responses in multiple separate feature maps (Treisman and Gelade 1980; Julesz
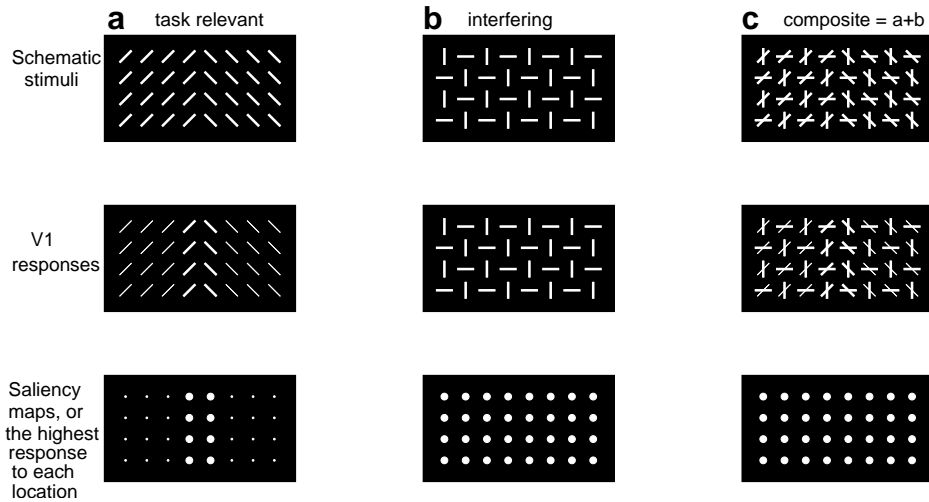
1981; Koch & Ullman, 1985; Wolfe, Case, Franzel 1989; Itti & Koch, 2000).

Figure 1: Psychophysical test of the V1 saliency hypothesis. **a, b, c**: schematics of texture stimuli (extending continuously in all directions beyond the portions shown), each followed by schematic illustrations of V1's responses and saliency maps, with bar thicknesses or disc sizes denoting response or salience levels. Every bar in **b**, or every texture border bar in **a**, has fewer iso-orientation neighbours to induce iso-orientation suppression, thus evoking less suppressed responses. The composite stimulus **c**, made by superposing **a** and **b**, is predicted to be difficult to segment, since the task irrelevant features from **b** interfere with the task relevant features from **a**, giving no saliency highlights to the texture border.

Of course, at least in primates, area V1 is the foundation of all cortical visual processing, and is also subject to substantial top-down influences itself, potentially mediating aspects of attentive processing that are important in later stages of visual computations. Nevertheless, the crippling restriction of the attentional bottleneck implies that this attentive processing is subservient to a surprising degree on bottom-up, pre-attentive, selection. The extent to which higher visual areas, such as V2 and beyond, contribute to pre-attentive selection and indeed meld pre-attentive and attentive influences is as yet unclear. We may certainly expect spatially mapped higher areas to execute a similar algorithm, detecting breakdowns in input homogeneity, but now breakdowns that are apparent in terms of the more sophisticated features to which their neurons are tuned (*eg* involving mid-level visual quantities such as surfaces; He & Nakayama 1992). The contribution of these areas may well be particularly manifest when selection on the basis of V1 features is least effective, and at a latency longer than that of pop out. Exploring this will be a fascinating task for the future.

3

References

[1] Allman J, Miezin F, McGuinness E. Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev. Neurosci.* 8:407-30 (1985).

[2] Gilbert C.D., Wiesel T.N., Clustered intrinsic connections in cat visual cortex. *J. Neurosci.* 3(5):1116-33 (1983)

[3] He, Z., Nakayama, K., Surfaces versus features in visual search. *Nature* **359**:231-3 (1992).

[4] Itti L., Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40(10-12):1489-506, (2000).

[5] Julesz B. (1981) Textons, the elements of texture perception, and their interactions. *Nature* 290(5802):91-7.

[6] Knierim JJ., Van Essen DC, Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* 67(4): 961-80 (1992)

[7] Koch C., Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4(4): 219-27 (1985).

[8] Li Z. Visual segmentation by contextual influences via intracortical interactions in primary visual cortex. *Network: Computation and neural systems* 10(2):187-212, (1999a).

[9] Li Z. Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proc. Natl Acad. Sci USA,* 96(18):10530-5. (1999b)

[10] Li Zhaoping 2002, "A saliency map in primary visual cortex " *Trends in Cognitive Sciences* Vol 6. No.1. Jan. 2002, page 9-16

[11] Neisser, U (1967) *Cognitive Psychology*. New York: Appleton-Century-Crofts.

[12] Nothdurft HC, Gallant JL, Van Essen DC. Response modulation by texture surround in primate area V1: correlates of "popout" under anesthesia. *Vis. Neurosci.* 16, 15-34 (1999).

[13] Rockland KS., Lund JS., Intrinsic laminar lattice connections in primate visual cortex. *J. Comp. Neurol.* 216(3):303-18 (1983).

[14] Sillito AM, Grieve KL, Jones HE, Cudeiro J, Davis J. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378, 492-496 (1995).

[15] Sziklai G (1956) Some studies in the speed of visual perception *IEEE Transactions on Information Theory* Vol. 2, Issue 3, p. 125-128

[16] Tehovnik EJ, Slocum WM, Schiller PH. Saccadic eye movements evoked by microstimulation of striate cortex. *Eur J. Neurosci.* 17(4):870-8 (2003).

[17] Treisman A. M., Gelade G. A feature-integration theory of attention. *Cognit Psychol.* 12(1), 97-136, (1980).

[18] Wachtler T, Sejnowski TJ, Albright TD. (2003) Representation of color stimuli in awake macaque primary visual cortex. *Neuron* 37(4):681-91.

[19] Wolfe J.M., Cave K.R., Franzel S. L. Guided search: an alternative to the feature integration model for visual search. *J. Experimental Psychol.* 15, 419-433, (1989).

[20] Zhaoping L. (2005) The primary visual cortex creates a bottom-up saliency map. In *Neurobiology of Attention* Eds L. Itti, G. Rees and J.K. Tsotsos, Elsevier, 2005, Chapter 93, page 570-575

[21] Zhaoping L. and May K.A. (2004) Irrelevance of feature maps for bottom up visual saliency in segmentation and search tasks. Program No. 20.1 *2004 Abstract Viewer/Itinerary Planner*, Washington D. C., Society for Neuroscience, 2004. Online.