

A model of global saliency: Selecting a target object among distractors.

Frank van der Velde¹, Gwendid T. van der Voort van der Kleij¹,
Pascal Haazebroek¹, Marc de Kamps²

¹*Cognitive Psychology Unit, Leiden University Leiden, The Netherlands*

²*Robotics and Embedded Systems, Technische Universität München, Germany*

Corresponding address: F. van der Velde, vdvelde@fsw.leidenuniv.nl

Abstract

We present a model of visual spatial selection (dorsal pathway). The model consists of a number of interacting spatial maps. The model selects the location of a single object (target) among a number of similar distractor objects. The target and distractors differ in shape or color. The target and the distractors are salient objects. Each one stands out on a uniform local background, which results in local “pop-out”. The location of the target object is selected due to the identity selection of either the target or the distractor (ventral pathway). Feedback activation based on object identification interacts with the spatial maps of the model. This interaction results in the selection of the location of the target.

Keywords: Dorsal pathway, Feedback, Saliency, Spatial selection, Ventral pathway.

1. Introduction

An object on a uniform background (e.g., a red ball on a green lawn) is highly conspicuous. The saliency of the object is due to its contrast with the background. In the retina, neurons already respond vigorously to local discontinuity as given by a contour or change in color or shading [1]. In perceptual terms, the object “pops-out” from its background and attracts attention. This form of attention has been modelled in a bottom-up manner, in which the local activity induced by the object dominates a retinotopic “saliency” map, by means of a winner-take-all (WTA) process [2].

Local saliency of objects also occurs in the case of, say, a red ball (target) and a number of blue balls (distractors) on a green lawn. Again, each object, target and distractors alike, are salient with respect to their local background. In neural terms, each object creates a discontinuity that results in vigorous activity in neurons in the early stages of the visual cortex. For the receptive field of these neurons is so small, that they are activated by the contrast of the object and local background, not by the entire object display. In terms of perception, each object thus pops-out from its local background. In terms of a bottom-up saliency model, each object generates strong local activity in a saliency map. Yet, the target still draws attention, that is, the red ball also pops-out among the distractors. Saliency in this case is global, because the target is salient not only with respect of its local background but also with respect to the entire visual scene.

Global saliency occurs with a number of stimulus features like color, shape, orientation, and shading [1]. Global saliency is different from local saliency, because it is not based on local discontinuity. Here, we present a neural model of global saliency.

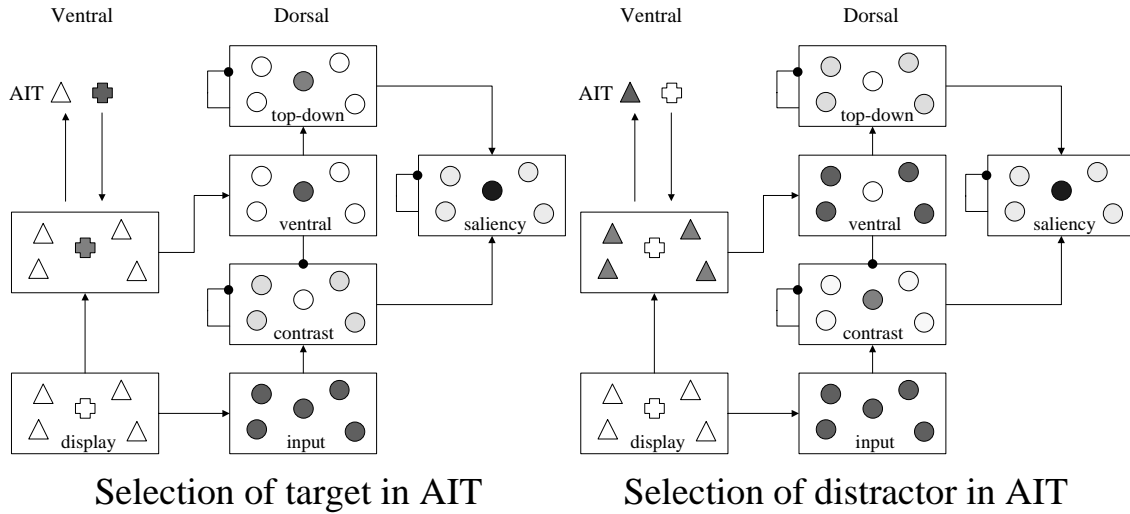


Figure 1. Model of spatial selection of a singleton (target) among distractors. Left: the identity of the target is selected in AIT (ventral pathway). The competition in the spatial maps of the model (dorsal pathway) results in the selection of the location of the target. Right: the identity of the distractor is selected in AIT (ventral pathway). The competition in the spatial maps of the model (dorsal pathway) again results in the selection of the location of the target.

2. A model of global saliency

Figure 1 illustrates a model for the selection of a target (e.g., cross) among distractors (e.g., triangles). It consists of two pathways: ventral and dorsal. The ventral pathway models object based attention as in [3]. Identity of an object is selected in ventral area AIT. It generates feedback activity, which interacts with stimulus activity in the ventral retinotopic areas. The result is the selection of activity related to the object’s location in these retinotopic areas. This selection (activation) is transmitted to the dorsal pathway.

In the dorsal pathway, the objects generate activation in an “input” retinotopic map. Activation is location related, not identity related. Each object is salient locally, so there is no difference in activation between the objects in the input map. The input map activates a “contrast” retinotopic map in a point-to-point manner (i.e., retinotopically). In the contrast map, WTA interaction [2] occurs between different spatial representations. The ventral pathway activates a “ventral” retinotopic map, in a point-to-point manner. The ventral map inhibits the representations in the contrast map in a point-to-point manner. The input and ventral map interact in the contrast map, so that the activation (“location”) that is not selected (enhanced) in the ventral map is selected. The ventral map also activates a “top-down” retinotopic map (point-to-point). In the top-down map, WTA interaction occurs between different spatial representations. Finally, the contrast and top-down map activate a “saliency” retinotopic map (point-to-point). In the saliency map, WTA interaction occurs between different spatial representations, as in [2].

Figure 1 (left) shows what happens when the target (cross) is selected in AIT. Its location is selected in the ventral map, and thus in the top-down map. The contrast map represents the location of the distractors (triangles), because the location of the cross is inhibited by the ventral map. Due to WTA, distractor representations are (more) reduced in the contrast map. As a result, the target’s location is most strongly activated in the saliency map. The target wins the WTA competition, and its location is selected.

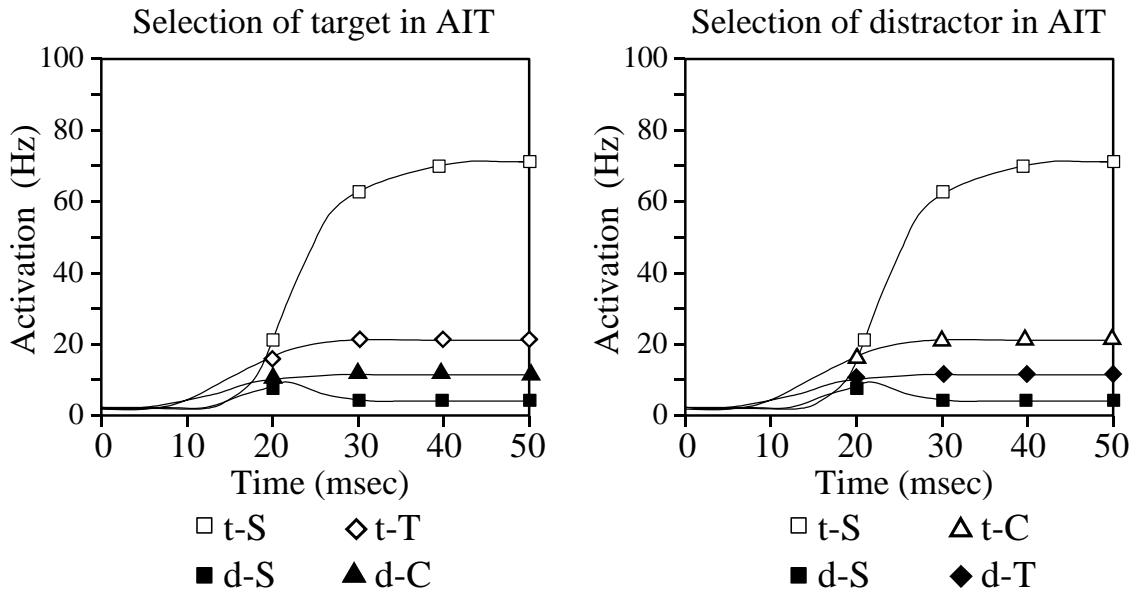


Figure 2. Left: activation in the spatial model when the target is selected in AIT. Right: activation in the spatial model when the distractor is selected in AIT. (t-S = target in saliency map, t-T = target in top-down map, t-C = target in contrast map, d-S = distractor in saliency map, d-C = distractor in contrast map, d-T = distractor in top-down map.)

Figure 1 (right) shows what happens when the distractor (triangle) is selected in AIT. The locations of the triangles are selected in the ventral map, and thus in the top-down map. But due to WTA, distractor representations are (more) reduced in the top-down map. The contrast map represents the location of the target (cross), because the distractor locations are inhibited by the ventral map. As a result, the location of the target (cross) is most strongly activated in the saliency map. The target wins the WTA competition, and its location is selected. The model operates likewise for, e.g., a red ball among blue balls.

3. Simulation of the model

Figure 2 presents the simulations of the model when the target (left) or the distractor (right) is selected in AIT. Simulations are based on population dynamics of spiking neurons, as in [3]. WTA competition is based on [2]. In both cases, the target is selected in the saliency map (t-S). Distractor activation in the saliency map is low in both cases (d-S). With target selection in AIT (left), target activity in the top-down map (t-T) is stronger than distractor activity in contrast map (d-C). With distractor selection in AIT (right), target activity in the contrast map (t-C) is stronger than distractor activity in top-down map (d-T). These differences in activation in favor of the target determine the selection of the target in the saliency map, in the manner as illustrated in figure 1.

References

- [1] S. Coren, L. M. Ward, J. T. Enns (2003). Sensation and perception. Wiley.
- [2] L. Itty, C. Koch (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Visual Search*, 40, 1489-1506.
- [3] F. van der Velde, M. de Kamps (2001). From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation, *J. Cognitive Neurosci.* 13, 479-491.