CLAM: Closed-loop attention model for visual search

Frank van der Velde[1], Marc de Kamps[2], Gwendid T. van der Voort van der Kleij[1]
[1]*Cognitive Psychology Unit, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands*
[2]*Robotics and Embedded Systems, Department of Informatics, Technische Universität München, Boltzmannstr. 3, D-85748 Garching bei München, Germany*
[1]{vdvelde, gvdvoort}@fsw.leidenuniv.nl, [2]kamps@in.tum.de

Abstract

We present the outline of a "closed-loop attention model" (CLAM), to model attentional phenomena such as visual search. CLAM consists of four 'parts': retinotopic areas, feature maps, spatial maps, and working memory. These parts form a closed-loop 'diamond' structure, with reciprocal connections. Visual search will result from the dynamic interactions in the diamond connection structure of CLAM. A target in working memory influences selection in the feature maps or the spatial maps, which then interact with the retinotopic areas. A target will be selected if there are not too many different distractors. Otherwise, a serial process will result, in which the interaction process in CLAM is limited to a restricted part of visual space.

Keywords: Attention, visual search, blackboard architectures, binding problem

1. Introduction

In previous work we modeled object-based attention in the temporal cortex [1,2] and visual working memory in the prefrontal cortex [3,4]. By integrating these models of object-based attention and working memory with a model of spatial attention in the parietal cortex, we aim to arrive at a "closed-loop attention model" (CLAM). The purpose of CLAM is to model attentional phenomena such as visual search.

Visual search consists of finding a target within a multitude of objects, or distractors [5]. Sometimes a target 'presents' itself (i.e., an 'oddball'), but often the target is stored in working memory. An influential model of visual search is the Feature Integration Theory (FIT) [6]. In FIT, visual search proceeds in two stages: a pre-attentive stage and an attentive stage. In the pre-attentive stage, elementary stimulus features (color, motion, size, orientation) are processed in parallel. In this stage, an oddball like a red object on a green background is detected forthwith. But more complex targets can only be found in the second attentive stage of the model. In contrast with the pre-attentive stage, the attentive stage operates in a serial manner, by selecting objects in sequence until the target is found. The objects are selected by spatial attention, that is, by selecting a particular location in the visual field.

In CLAM, the pre-attentive and attentive stages are integrated, in line with experimental findings that do not support a clear distinction between pre-attentive and attentive processing [5]. Furthermore, in contrast with FIT, selection can result from any of the processed stimulus features (color, shape, motion, location), including complex features such as whole shapes.

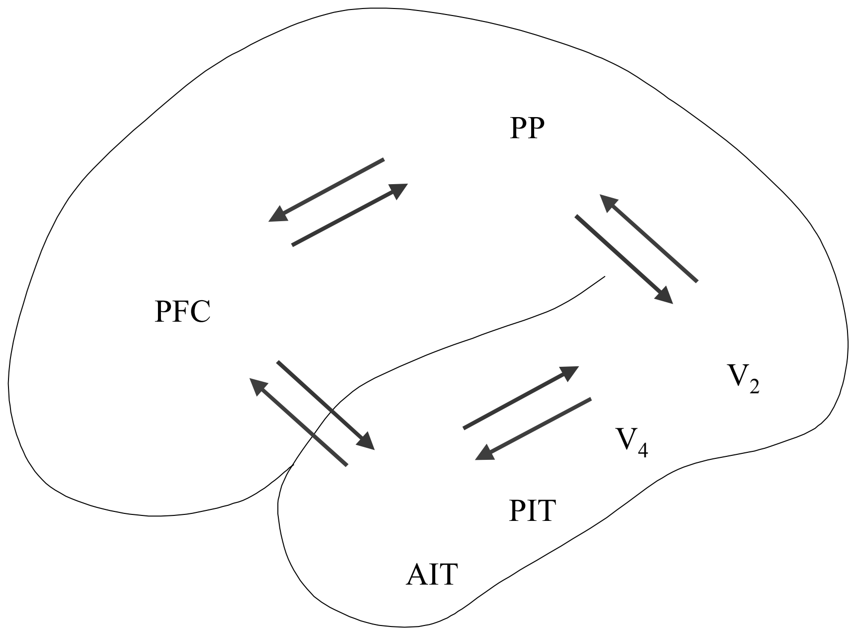

Figure 1. The overall connection structure of the Closed-Loop Attention Model. PFC = prefrontal cortex; AIT = anterior infero-temporal cortex; PIT = posterior infero-temporal cortex; PP = posterior parietal cortex.

Figure 1 illustrates the overall connection structure of CLAM. The model consists of four 'parts'. The first part consists of the (lower) retinotopic areas of the visual cortex (e.g., V2-PIT). The second part consists of the networks (or ‘feature maps’) in the ventral pathway that process object identity (e.g., shape, color). The third part consists of the networks (or ‘spatial maps’) in the dorsal pathway that process location information of objects in the visual field, and that transform this information into spatial coordinates for specific movements (e.g., eye, body, head, arm). The fourth part consists of visual working memory areas in the prefrontal cortex. The four parts are connected in a 'diamond' structure, with reciprocal connections. In this way, the diamond connection structure of CLAM forms a closed loop.

Figure 2 illustrates the functional structure of CLAM. Processing in CLAM starts in the retinotopic areas. The neurons in these areas have (relatively) small receptive fields and they typically encode conjunctions of elementary visual features. For instance, they encode for elementary conjunctions of shape (e.g., orientation) with color, or of conjunctions of shape with motion (such as an oriented bar moving in a particular direction). Because the areas are retinotopic, the neurons encode for location as well.

The ‘ventral’ and ‘dorsal’ pathways in CLAM emerge from the (lower) retinotopic areas. The ventral pathway transforms the retinotopic information into location-invariant feature information about object identity (shape, color). In figure 2, the ventral pathway processes the feature information (shape and color) of a display that consists of a dark (‘blue’) cross on the left and a light (‘yellow’) diamond on the right. The 'dorsal' pathway processes the spatial (location) information of the objects in this display. In CLAM, the ventral and dorsal pathway each consists of a combination of a feedforward network and a feedback network, which interact locally [1].

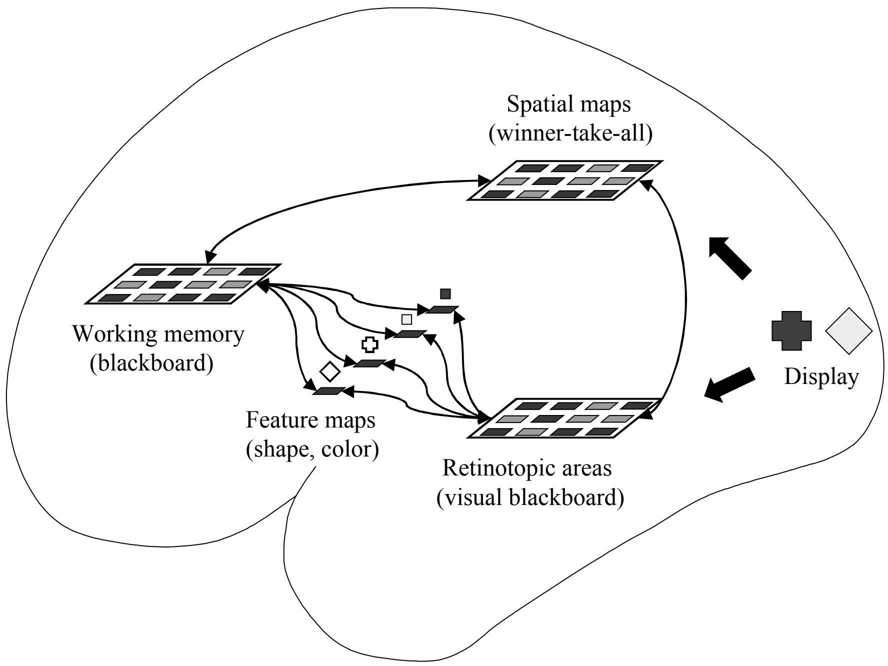

Figure 2. Functional structure of the Closed-Loop Attention Model.

Interaction between the ventral and dorsal pathway occurs in the retinotopic areas (e.g., V2-PIT). These areas function as a 'blackboard' [3] in which the features of an object (shape, color, motion, location) can be related or bound. The notion of a ‘blackboard’ derives from the fact that representations in these areas combine elementary feature (shape, color) information with location information. If one feature of an object (e.g., its shape or color) is selected as a cue (object-based selection), the other features of the object (including its location) can be selected as well by means of an interaction process in the blackboard (as illustrated below). Likewise, the selection of the location of an object (location-based selection) can be used to select the features (shape, color) of the object by means of the interaction within the blackboard.

The ventral and dorsal pathway in CLAM also project (feedforward) to the prefrontal cortex (PFC). In the PFC, the features of a target object (or objects) are stored in a working memory (WM) 'blackboard' [3,4]. The 'WM-blackboard' in PFC is similar in nature to the blackboard in visual cortex (e.g., on the level of retinotopic representation in PIT). It interacts with location-invariant feature representations (shape, color) that are either located in the ventral pathway or in the PFC itself (or perhaps both). It also interacts with location representations that are either located in the PFC or in the dorsal pathway (or both). The working memory blackboard is used to bind the features (shape, color, location) of an object stored in working memory. The working memory in PFC projects back to the ventral and dorsal pathway, through the representations for features and location.

2. Object-based selection in CLAM

Figure 3 illustrates the process of object-based selection in CLAM. A feature of a target object (the shape of the cross) is stored in the WM-blackboard. For instance, the cross (without a color) was presented earlier on the center of a display. Then, after a delay period, a display of two objects is presented, and the subject has to select the other features of the cued object (the cross).

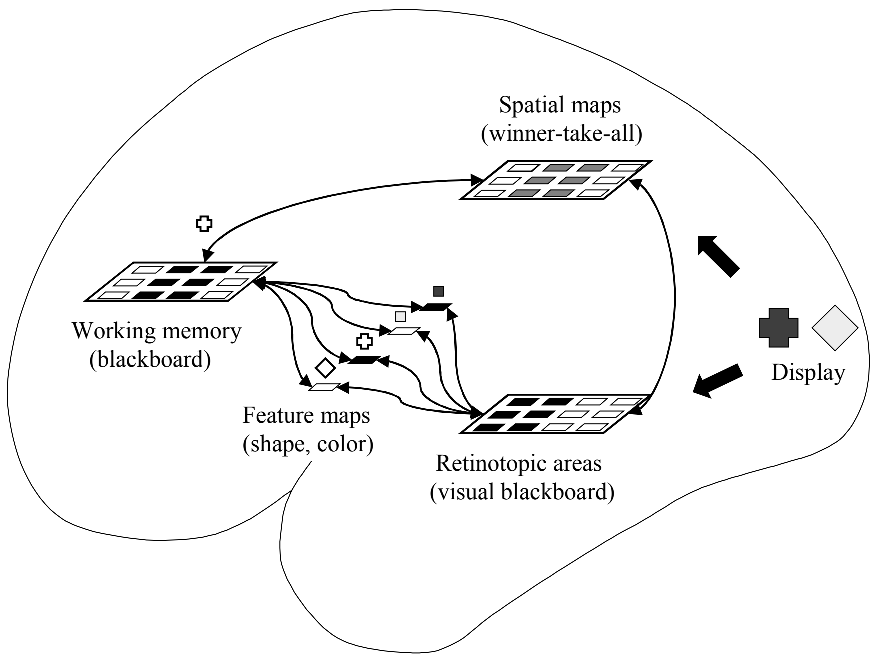

Figure 3. An object-cue in working memory initiates object selection in the Closed-Loop Attention Model.

In CLAM, the selection of the shape (or the color) of a target object by a cue results in enhanced activation on the location of the target in the visual blackboard (V2-PIT). This enhanced activation results from the interaction between the feedforward network and the feedback network in the ventral pathway. The feedforward network processes the identity of the objects in the display (shape, color). The feedback network in the ventral pathway carries the information of the cue back to the retinotopic areas (the visual blackboard). The cue-related activation in the feedback network is initiated by the information stored in the WM-blackboard.

The activations in the feedforward network and the feedback network match on the location of the target (cross) in the visual blackboard (corresponding to the location of the cross in the display), which results in the enhancement of all activation in the visual blackboard that is related with the cross [1]. In this way, the color of the cross can be selected in the ventral pathway [2]. Likewise, the location of the cross in the visual display can be selected in the dorsal pathway [1]. Notice that the WM-blackboard would also activate the dorsal pathway in relation with the location of the cue as it was initially presented. However, this location does not correspond to a location of an object in the display, and it is not enhanced by the selection of the target location in the ventral pathway. Hence, the location stored in the WM-blackboard would not be selected in the spatial maps of the dorsal pathway by a winner-take-all process.

3. Spatial-based selection in CLAM

Figure 4 illustrates the process of spatial-based selection in CLAM.

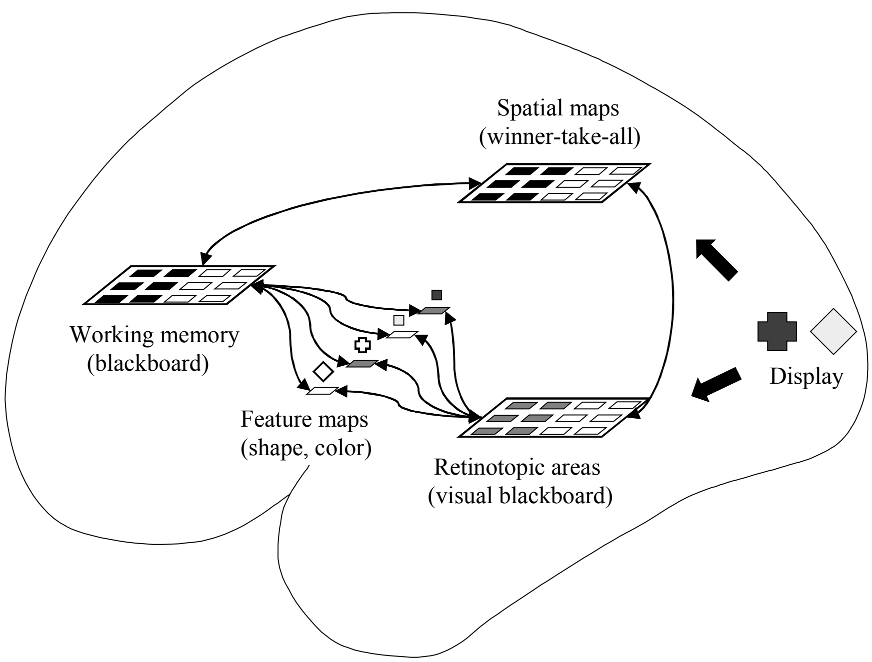

Figure 4. A spatial cue in working memory initiates object selection in the Closed-Loop Attention Model.

A spatial cue (without any identifiable shape) can be stored in the WM-blackboard. This will result in an enhanced activation in the dorsal pathway that selects the location of one object (target) in a visual display. In turn, the selection of a location in the dorsal pathway will enhance activation on that location in the retinotopic areas (V2-PIT), which results in the selection of the shape and the color of the object on that location in the ventral pathway, in line with the notion of spatial attention.

4. Visual search in CLAM

Visual search is based on the dynamic interactions in the connection structure of CLAM. A cue (shape, color) in PFC influences the ventral pathway and the dorsal pathway, which results in the interaction in the visual blackboard described above. If the cued target is unique, this interaction will result in the selection of the activation related with the target, which produces the selection of all other features of the target (in the ventral or dorsal pathway). However, all objects with features similar to the cue (e.g., color) will be selected in this way. In CLAM, this will result in a winner-take-all process in the spatial map. The selected location in that process will enhance the activation in the visual blackboard, which selects all features on that location. This process continues until the selected features match the features in WM.

The selection of the target will not succeed if there are too many different objects in the visual blackboard [4]. In this case, and in the case when a target consists of a conjunction of features (e.g., a particular shape-color combination), a serial process is needed to select the target. In CLAM, this serial process is location-based (as in FIT). It consists of pre-selecting a spatial region in the visual blackboard, in which the interaction process described above occurs. This serial process continues until the target is selected.

Visual search in CLAM is influenced by object saliency, e.g., due to the onset of a (new) object. Saliency influences the winner-take-all process in the spatial map. This causes an interaction with the search process in CLAM, which can result in the (temporal) selection of the location of the salient object over the location of cued target.

References

[1] Van der Velde, F. & de Kamps, M. (2001). From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation. *Journal of Cognitive Neuroscience. 13 (4)*, 479-491.

[2] De Kamps M. & van der Velde F (2001). Using a recurrent network to bind form, color and position into a unified percept. *Neurocomputing, 38-40,* 523-528.

[3] Van der Velde, F. & De Kamps, M. (2003). A model of visual working memory in PFC. *Neurocomputing*, *52-54*, 419-424.

[4] Van der Voort van der Kleij, G. T., de Kamps, M. & van der Velde, F. (2003). A neural model of binding and capacity in visual working memory. *Lecture Notes in Computer Science*, *2714*, 771-778.

[5] Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences, 7 (2)*, 70-76.

[6] Treisman, A, & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.