A Neural Model of Binding and Capacity in Visual Working Memory

Gwendid T. van der Voort van der Kleij[1], Marc de Kamps[2], and Frank van der Velde[1]

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

Abstract. The number of objects that can be maintained in visual working memory without interference is limited. We present simulations of a model of visual working memory in ventral prefrontal cortex that has this constraint as well. One layer in ventral PFC constitutes a 'blackboard' representation of all objects in memory. These representations are used to bind the features (shape, color, location) of the objects. If there are too many objects, their representations will interfere in the blackboard and therefore the quality of these representations will degrade. Consequently, it becomes harder to bind the features for any object maintained in memory, which reduces the capacity of working memory.

1 Introduction

Recent investigations [1] have shown that humans have the ability to maintain a number of visual objects in visual working memory. A remarkable characteristic of this finding is that the number of objects that can be maintained in working memory without interference (i.e., loss of information) is limited (to about four), but the number of object features (e.g., shape, color, location, motion, etc.) is unlimited for each of the objects. A model of visual working memory in prefrontal cortex (PFC) has been presented that can explain this characteristic [2]. Basically, this model is characterized by a 'blackboard' that can link different 'processors' to one another. The processors in this case are networks for feature identification (shape, color, location). One layer in ventral PFC functions as the blackboard, containing representations that consist of conjunctions of (partial) 'identity' (shape, color) information and location information. This blackboard serves to bind the information processed in each of the specialized feature networks. Objects in working memory are stored in the blackboard. When too many objects are put in working memory, their representations in the blackboard interfere. Consequently, an object's representation in the blackboard muddles and the capacity of the blackboard to bind the features of an object degrades.

O. Kaynak et al. (Eds.): ICANN/ICONIP 2003, LNCS 2714, pp. 771–778, 2003.

After getting deeper into this model of visual working memory, we present simulations that confirm our expectations that the model is limited in the number of visual objects that it can maintain without interference.

2 Blackboard Architecture of Visual Working Memory in PFC

Our model of visual working memory in PFC is based on a neural blackboard architecture that is used in a simulation of object-based attention in the visual cortex [3]. We assume that the neural blackboard architecture is located in the ventral prefrontal cortex (V-PFC) [2]. This is in line with human neuroimaging studies and recent monkey studies [4]. Activation in V-PFC is sustained (reverberating) activation, characteristic of working memory activation in the cortex.

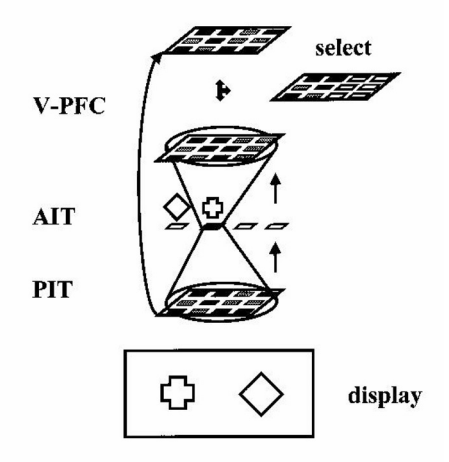

Fig. 1. A blackboard architecture in prefrontal cortex (PFC). PIT = posterior infero-temporal cortex; AIT = anterior infero-temporal cortex; V-PFC = ventral prefrontal cortex.

In the model (figure 1), the ventral prefrontal cortex (V-PFC) has a layered structure with representations similar to the representations in the visual (temporal) cortex. First, the posterior infero-temporal cortex (PIT) connects to one of the layers in V-PFC (for the purpose of illustration: the top layer in figure 1). As in PIT itself, the representations in this layer of V-PFC consist of conjunctions of location and (partial) identity (object-feature) representations (shape, color). In turn, another layer of V-PFC (the bottom layer in figure 1) is connected to the higher-level areas in the visual cortex, in which location and (location-invariant) object identity information are processed and represented (in figure 1 illustrated for the anterior infero-temporal cortex (AIT), where the

shape of an object is processed and represented). These connections are similar to the connections of the feedback network of the visual cortex in [3]. They have a 'fan-out' structure, which means that they connect to all possible representations that are selective for an activated feature (on every possible position). As a result, the representations in the bottom layer of V-PFC consist of distributed identity representations. The bottom and top layer of V-PFC interact in a manner similar to the interaction between the feedforward and feedback networks of the visual cortex in [3], using similar microcircuits. This interaction results in the selective activation of a third layer in V-PFC (the 'select' layer in figure 1). In particular, in the select layer there is activation on locations in which there is a substantial match in activation between the top and bottom layer of V-PFC.

Figure 1 illustrates the selection process in the V-PFC model. In figure 1, two objects are processed in the visual cortex, and their PIT representations also activate the representations in the top layer of V-PFC. The activation of one of the objects (the cross) is selected (attended) in AIT (e.g., due to a competition between both figures in AIT). This identity activation of the cross in AIT activates the bottom layer of V-PFC. As a result, the interaction between the top and bottom layer activate the representations in the select layer that are selective for the features (e.g., shape, color, position) of the cross. The activation in the select layer can be used to activate the other features of the cross [3,5].

2.1 Feature Binding in Working Memory

The nature of the representations in V-PFC and the connections with the higher-level areas in the visual cortex produces the behavioral effects described before. The blackboard architecture of V-PFC results in a binding of the feature representations of the objects maintained in memory. Therefore, the features of an object can be retrieved (selected) in working memory as long as the representations of the objects stored in V-PFC do not interfere. However, when too many objects are present in a display, their representations in V-PFC will interfere, which results in loss of information. As more objects are present in a display, the amount of interference increases, and it can be expected that the quality of the representation of an object in V-PFC becomes less. As a consequence, it becomes harder to correctly bind the feature representations of the object that are maintained in memory. V-PFC might end up binding wrong feature representations for an object that is attended to. We carried out simulations to see whether our model of the visual working memory shows this behavior.

3 Simulations

For the simulations we used the same neural network model of (the ventral pathway in) the visual cortex that is used in the simulation of object-based attention in the visual cortex [3]. It basically consists of a feedforward network that includes the areas V1, V2, V4, PIT and AIT, and of a feedback network that carries information about the identity of the object to the lower areas

in the visual cortex (V1 - PIT). The model shares the basic architecture and characteristics (i.e., the nature of the representations) of the visual cortex. For the purpose of our simulations, we trained 5 feedforward neural networks to identify 9 different objects on 9 possible positions (using backpropagation). After a feedforward neural network had successfully learnt this task, its corresponding feedback network was trained as well (using Hebbian learning) [3]. This resulted in having 5 instances of the visual cortex model, with each instance having slightly different connection weights between its layers.

The layers of the visual working memory were subsequently simulated as follows. The activation in the top layer of V-PFC is simulated as a copy of the activation in PIT after a display is processed feedforwardly through the visual cortex [3]. This is done because the representations in this layer of V-PFC are similar to the representations in PIT. For reasons of simplicity, the bottom layer of V-PFC, which is connected to many higher-level areas in the visual cortex, is simulated being connected to just one of these areas, AIT. The connections from AIT to this layer are similar to the connections between AIT and PIT in the feedback network of the visual cortex [3]. These connections are therefore copied from a trained feedback network and the representation in this layer equals the representation in PIT in the feedback neural network.

During the simulations, displays consisting of N (different) objects, with N ranging from 2 to 9, are presented to V1. For each N, 180 random displays are presented to each instance of the model. Objects in a display are placed on separate, non-overlapping, positions. Let us see what happens in our model after presentation of a single display (i.e., one trial). First, the visual cortex processes the display. The feedforward neural network gradually transforms the retinotopic information in the primary visual cortex into identity-based information. The representations of the objects in PIT also activate the representations in the top layer of V-PFC, that receives its information from PIT. This layer of V-PFC stores the objects in the display. Now suppose that one of the represented objects is attended to (selected). The attended object activates its shape representation in AIT, and consequently, all representations in the bottom layer of V-PFC that are selective for the shape of the object. The question now arises whether it is possible to bind the shape of the attended object with its other features (e.g., its location), despite the fact that $N-1$ other objects are also present in the display. If the representation of the attended object in the top layer of V-PFC is still intact (i.e., is not severely affected by the representations of other objects), the interaction between the top and bottom layer can activate the representation in the select layer that is selective for the attended object. This implies that this representation should be activated in the select layer on a position that corresponds to the location of the attended object in the display. But, in the case that the N representations of the objects in the top layer interfere too much, and make each other's representations 'fuzzy', the interaction between the top and bottom layer cannot uniquely activate the representation that is selective for the attended object anymore. Instead, it might wrongly activate a representation that (originally) is selective for another object. Feature binding

of the selected object then fails. The chance of this happening will likely rise as the number of objects represented in the top layer of V-PFC increases.

For example, if a cross and a diamond are presented in a display (figure 1), the cross on the left and the diamond on the right, then this display will be represented in the top layer of V-PFC. Selecting the cross in AIT subsequently activates the distributed representations of the shape of the cross at any possible position in the bottom layer of V-PFC. By means of the interaction between the top and the bottom layer, the representation of the cross on the left position in the select layer will be activated. But, in the case that the representation of the cross and the representation of the diamond in the top layer are interfering too much, the selection of the cross in AIT could result in the incorrect activation of a representation on the right in the select layer. Let us see how our model of visual working memory behaved.

4 Results

In our model of visual working memory, the representation in the encode layer embodies the match between the representation in the top and the bottom layer of V-PFC.

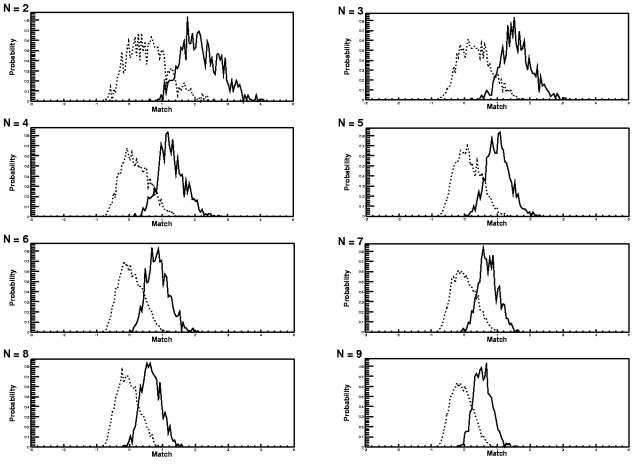

Fig. 2. Probability distribution of match (i.e., standardized positive covariance per position) for positions of attended objects (solid line) and for positions of unattended objects (dashed line) in the top layer in V-PFC. Y-axis: probability. X-axis: match, from negative (left) to positive (right).

The artificial neurons can have activation values in the range −1 to 1. Positive and negative activation can be regarded as activity of separate populations

of neurons [6]. Thus, negative activation in the bottom layer and negative activation in the top layer is also a match. Therefore, we simulated the interaction between the top and the bottom layer of V-PFC by computing the covariance between them. Note that these covariance values offer two kinds of information; the match (positive covariance) and the mismatch (negative covariance). After every presentation of a display with N objects, the positive covariance for every possible position of an object in the blackboard (top) layer was summed and subsequently standardized by the average positive covariance per position during that trial. The same was done for the negative covariance. We will further refer to this standardized positive and negative covariance as the match and mismatch respectively.

It may be clear that within every trial, one position in the top (and select) layer corresponds to the position of the attended object in the display, and $N-1$ positions in these layers correspond to positions of objects in the display that are unattended. The rest of the positions in the top and select layer $(9-N)$ correspond to locations in the display where no object was presented.

Figure 2 shows the probability distribution over several amounts of match for positions in the top layer of attended objects and unattended objects separately. For each number of objects in working memory, data of all 5 instances of the neural network model are averaged over all relevant trials. Note that for successful binding to occur, the match should be high on the position of the attended object and low on positions of unattended objects (as the mismatch should be respectively low and high). Only then the position of the attended object can be clearly distinguished from the positions of unattended objects in terms of match. As can be seen in the figure, this is the case if the number of objects held in working memory is low.

Figure 3 shows the probability distribution over several amounts of mismatch for positions in the top layer of attended objects and unattended objects separately. Again, for each number of objects in working memory, data of all 5 instances of the neural network model are averaged over all relevant trials. Note that for successful binding to occur, the mismatch should be low on the position of the attended object and high on positions of unattended objects (as the match should be respectively high and low). Only then the position of the attended object can be clearly distinguished from the positions of unattended objects in terms of mismatch. Again, as can be seen in the figure, this is the case if the number of objects held in working memory is low.

However, figures 2 and 3 show that the probability distribution of match and mismatch for the positions of attended objects and for the positions of unattended objects start to overlap more and more as the number of objects in working memory increases. This means that the position of the attended object cannot be reliably selected on the basis of positive covariance. As the load on the visual working memory gets higher, positions of unattended objects will more frequently be selected instead. In other words, the binding process starts to break down.

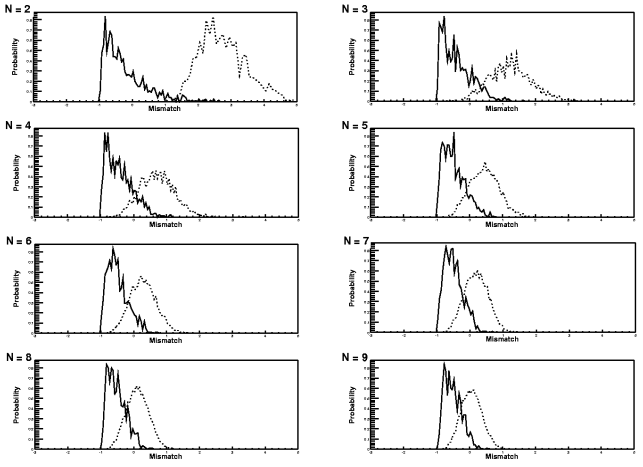

Fig. 3. Probability distribution of mismatch (i.e., standardized negative covariance per position) for positions of attended objects (solid line) and for positions of unattended objects (dashed line) in the top layer in V-PFC. Y-axis: probability. X-axis: mismatch, from negative (left) to positive (right).

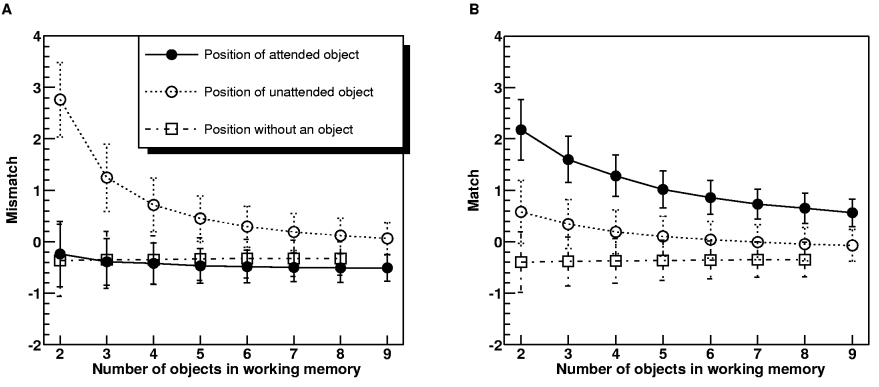

Fig. 4. (A) Mismatch (mean and rms) on positions of attended objects (solid line), on positions of unattended objects (dot-dot line) and on positions without an object (dash-dot line) in the top layer in V-PFC. (B) Idem, but then for match (i.e., standardized positive covariance per position).

The mean amount of match for positions of attended objects, positions of unattended objects and positions with no object is presented in figure 4B together with its root mean square (rms). Picking the position of the attended object instead of a position of an unattended or empty position on the basis of match information clearly becomes very hard as the number of objects in working memory increases. Does mismatch information enable us to point out

the right position of an attended object when the number of objects stored in memory increase? The answer is given in figure 4A, and appears to be negative. The distinction between attended and unattended objects gets lost here as well. Filling up the working memory makes the level of mismatch that can be detected in the top layer on the position of the attended object more and more similar to the level of mismatch on other positions. Thus, based on mismatch information, binding begins to fail as well.

5 Discussion

The simulations point out that the model of visual working memory that we presented is limited in the number of objects that it can maintain in memory without interference (i.e., loss of information). Our model cannot successfully bind the feature(s) of the attended object anymore as it gets loaded with more objects. This is in accordance with findings about visual working memory [1]. However, when exactly the limit in visual working memory is reached will depend on other factors as well, like the level of alertness and the contrast of the objects with the background. We predict that this limit is also partly dependent on the distance between objects in a display. Objects that are close to each other activate more common neurons in the top layer of V-PFC than objects that are far from each other. More overlap between representations of objects in the top layer of V-PFC leads to more interference and thus enhances the chance of binding the wrong features for an attended object.

References

1. Vogel, E.K., Woodman, G.F., Luck, S.J.: Storage of features, conjunctions, and objects in visual working memory, Journal of Exp. Psychol.: HPP **27** (2001) 92–114
2. Van der Velde, F., de Kamps, M.: A model of visual working memory in PFC, Neurocomputing (2003) (in press)
3. Van der Velde, F., de Kamps, M.: From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation, J. Cognitive Neurosci. **13** (4) (2001) 479–491
4. Duncan, J.: An adaptive coding model of neural function in prefrontal cortex, Nature Rev. Neurosci. **2** (11) (2001) 820–829
5. De Kamps, M., van der Velde, F.: Using a recurrent network to bind form, color and position into a unified percept. Neurocomputing **38-40** (2001) 523–528
6. De Kamps M., van der Velde, F.: From artificial neural networks to spiking neuron populations and back again. Neural Networks **14** (2001) 941–953