HAL
open science

Spatial Knowledge Transfer Between Models of Hippocampus and Associative Cortex

Nicolas P. Rougier, Frédéric Alexandre

To cite this version:

Nicolas P. Rougier, Frédéric Alexandre. Spatial Knowledge Transfer Between Models of Hippocampus and Associative Cortex. International Joint Conference on Neural Networtks, 1999, Washington, D. C., 5 p. inria-00098757

HAL Id: inria-00098757

https://inria.hal.science/inria-00098757v1

Submitted on 26 Sep 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spatial Knowledge Transfer Between Models of Hippocampus and Associative Cortex

Nicolas Rougier, rougier@loria.fr and Frédéric Alexandre, falex@loria.fr,

LORIA, BP 239, 54506 Vandoeuvre Cedex, France

Abstract—*A* typical navigation task requires both declarative and non declarative memory for respectively locus memorization and action selection. These dual memories are believed to be achieved in the central nervous sytems by the hippocampus and the cortex. To the light of the functionning of each neuronal system, this paper proposes a model of interaction between these structures. This model is illustrated through a simple example of navigation task.

I. Introduction

Rodent spatial navigation abilities have been extensively studied during this century, and some 20 years ago, place cells were predicted and actually found in the hippocampus of rats [10]. Since that time, many computational models have been designed and solely based on this main feature [6], allowing efficient navigation in miscellaneous environments via the use of "cognitive maps". In the same time, some other models were mainly concerned with possible interactions between neocortex and hippocampus, especially investigating the way memory can be transfered from hippocampus to cortex [8]. Today, declarative memory is attributed to the hippocampal structure and procedural memory is attributed to the neocortex.

In the framework of autonomous robot navigation we want to investigate the ways of using the declarative memory stored by the hippocampus. This declarative memory is indeed required for several cognitive tasks and cannot be reduced to spatial memory only. In the same way, a model of cortex cannot be reduced to a spatial processing unit. Hence, we propose here an architecture whose role is to link events or facts (whatever their meaning). We assume that each state, which corresponds to a cortical pattern of activation, can be linked to the next state and indexed by the action which allowed the transition. We then need a basic model of primary cortex able to process external flow of inputs, a model of hippocampus able to store cortex pattern of activation and recall them when needed, and finally a model of associative cortex able to store indexed links between different states.

II. The hippocampus for declarative learning

The hippocampus system of rodents, monkeys and humans has been extensively studied during the past decades, leading to the design of several theories concerning its role in the overall brain functionning. Hypotheses concerning the role of the hippocampus in the rodent brain are strongly linked to navigation aspects since [10], [9] found place cells in the freely moving rat in this neuronal structure. Actually, it has been observed that some pyramidal cells in regions CA3 and CA1 in the hippocampus possess the following properties :

- They rapidly fire only when the rat's head is in a particular region of the environment
- This region is cell-specific for a given environment and is fixed for each individual cell over a long period of time
- These cells remain silent outside their "fire field"

These observations led to the theory that the rat hippocampus would be able to construct some cognitive map of the environment [11] allowing the rat to go from one place to another. In fact, this theory postulates that the rat is able to construct some allocentric model of the environment using environmental cues. Moreover, [6] designed an hippocampus model for spatial navigation which could also take the dead recknoning feature into account, i.e. the use of path shortcuts that were never used before. This model uses a compass-like system where distance and bearing are stored within the hippocampus, allowing this way to compute new paths. These models indeed endow the hippocampus with spatial memorization and computation abilities.

Global studies bring out quite different theories regarding the role of the hippocampus in the brain processes. Early back in 1957, the study of the famous patient H.M. [13] pointed out the essential role played by the hippocampus in some memory tasks. The temporally retrograde amnesia induced by hippocampus impairment suggests that this structure plays an important role in a precise kind of memory tasks : the

episodic declarative memory. As related in [5], [4] proposed a theory about the crucial role of the hippocampus in memory consolidation. He suggested that the hippocampal system stores experiences and plays them back to the neocortex, where categorization would be performed. The very idea that **hippocampus is the neocortex teacher** has been now widely adopted by most researchers in this domain : [5], [12], [8], [7] and [14]. Indeed, the anatomy of the brain shows that the hippocampus is linked to the whole neocortex via the entorhinal cortex and is thus able to take a complete snapshot of its content and snapshoting the neocortex is equivalent to explicitly memorizing an experience, because all necessary data are present within this structure. Moreover, the mechanism of LTP (Long Term Potentiation) discovered in the hippocampal structure allows for rapid formation of memory, in opposition to the longer cortical learning mechanisms. Despite fine differences, recent neurobiological models converge toward that role of the hippocampus in the formation of neocortical representations.

III. The cortex for non declarative learning

A. Overview

The cortex can be considered as a set of functional neural areas including two different types :

- monomodal areas : They mainly regroup the sensory (e.g. visual) and the effector (e.g. motor) areas
- polymodal areas : They are the associative areas linking at least two other (monomodal or polymodal) areas

Some of these areas are connected to extra-cortical structures and are also interconnected, allowing the propagation of activation from one area to some others. The origin of sensory input and the pattern of connectivity determine the nature of information processed within monomodal and polymodal areas. Both kinds of areas build with learning a statistical representation of information correlation leading to such tasks as speaking, memorizing friend faces or moving hands, deeply embedded within connections between these areas. In this framework, learning processes within the cortex are related to a large set of slow local neuronal activity-dependent modifications. These modifications correspond to non declarative learning because they correspond to a slow process of extraction of functionnal relationship between monomodal and polymodal dimensions [1].

B. The cortical neural assemblies

Several studies of the human cortex lead to a better understanding of mechanisms involved in cognitive tasks completion. Particularly, some of them are carried out at the unit level. The cortex is indeed made of several elementary units : the cortical columns which are assemblies of neurons and have been extensively studied in [2]. From a functionnal point of view these neural assemblies can be seen as a cellular automaton which can be activated at four different levels of activity:

- Level E_{-1} : **inhibited** state
- Level E_0 : **neutral** state
- Level E_1 : **calling** state
- Level E_2 : **excitated** state

While inhibited, neutral and excitated states are quite familiar for usual neuronal units, the calling state is characteristic of the cortical units. Roughly, it can be said that the excitation of the cortical column requires two weighted preconditions :

- Excitation from thalamic afferences
- Excitation from cortical afferences

In most cases, the activation of both afferences are required to excite the column. Moreover, the cortical afferences, when not combined with thalamic afferences are spread to all neighbours (modulo link weights). Indeed, the state E_1 is seen as a calling state whose signification is a requirement for the column to become excited (cf. fig. 1).

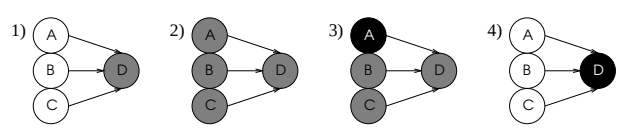

Fig. 1. Functionnal view of columnar activations : **1.** At time t, the perceptive unit D is calling to be excited. **2.** At time $t+1$, call has reached units A, B and C. **3.** At time $t+2$, the context allows action A to be performed and the unit A becomes excited (because of thalamic afferences). **4.** At time $t+3$, the context has been changed because of the previous action, and then unit D becomes excited. Here, learning occurs, and unit D will little by little preferentially call unit A.

Cortical areas are indeed made of several columns (connected to each others via lateral links) which receive the same kind of information. According to the weight of connections, each information pattern received by the area induces excitation of a small subset

of those columns while the others are inhibited by lateral connections. Learning indeed enhances this mechanism up to the point where columns are completely uncoupled. This splitting mechanism allows a strong contrasted activity within each cortical area.

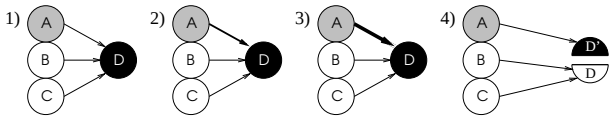

Fig. 2. When a sequence of activation is repeated along time, the column is able to split itself in order for the created column to be tuned to this specific sequence.

The exact conditions of splitting may greatly vary from one area to the other and mainly depend on the location of the area the column belongs to.

IV. Experiments

A. The task

The task to be completed is quite simple : a simulated robot is placed in a square world where each locus is perceived via 4 colored shapes (cf. fig. 3) and the robot is able to move with absolute motor action (up/down/left/right). The world is "frozen" : the perception of the external world is unchanged as long as no effector action is performed.

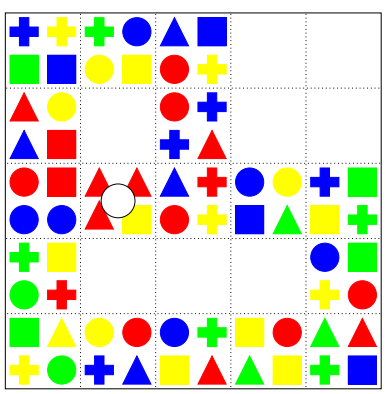

Fig. 3. Environment sample : Each locus is characterized by 4 colored shapes (red/green/blue/yellow and square/triangle/circle/cross).

In the first period, the simulated robot is controled and moved all over the maze for him to learn the map (i.e. learning the links between two loci).

In the second period, the robot is "given" a perception corresponding to a locus (anywhere in the map) and is expected to go to this place using a non-random path (it is not required that the shortest path is used).

B. Global model architecture

The model is based on a perception-action loop where perception is a 4-areas retina and action is a 4-actions motor system (cf. fig. 4).

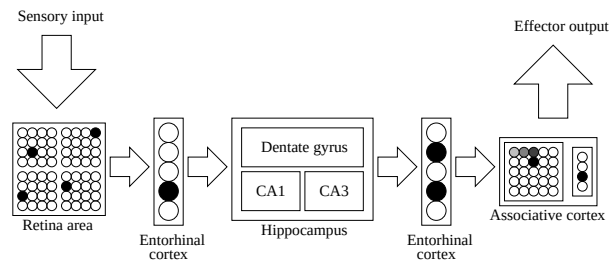

Fig. 4. Flow of information within the model.

In the first period, 4 shapes, corresponding to the current location, are simultaneously perceived by the retina. Then, retina activation is propagated to the entorhinal cortex (EC). This activation is then learned at the level of the hippocampal structure and remapped onto the entorhinal structure. Since the previous location was perceived few steps before, the corresponding activation traces are also presents within the entorhinal cortex. Then, the associative cortex has to learn the triplet(locus L_1 - Action A - locus L_2). Once the entire map is learned, the model has indeed encoded every possible relation between all places within some units. Then, the call property of column will allow to retrieve any path using the water-flood algorithm.

V. Model of hippocampus

A. Structure

The hippocampus model is composed of four distinct structures:

- Dentate gyrus (DG) : This structure is composed of units which are linked to a random set of few units (between 2 and 4) in the entorhinal cortex and to one unit in CA3.
- CA1 : This structure is composed of units which are linked to a random set of few units of the dentate gyrus.
- CA3 : This structure is composed of units which are reciprocally linked to a random set of few units in the dentate gyrus and to all other unit in structure CA3. Each unit is also linked to its corresponding unit in the CA1 structure.
- Septum : This structure is composed of a single unit which is linked to all units in structures CA1 and CA3.

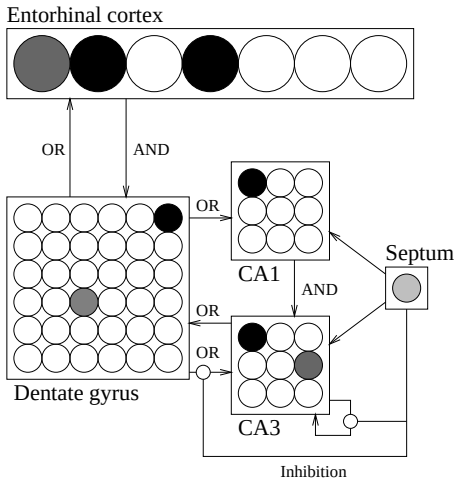

Fig. 5. The hippocampal structure.

B. Activity computation

- DG units: These units perform a logical AND function onto the EC units they are linked to, hence allowing some overlapping pattern of the EC to be less overlapped within the DG. They may be also directly activated by CA3 units.
- CA1 units : These units perform a logical OR function onto the DG units they are linked to, hence giving a pattern reference for CA3.
- CA3 units : These units perform a logical OR function onto the DG units they are linked to and onto the other CA3 units. They can also be directly activated or inhibited by the corresponding CA1 units when septum is excitated enough.
- Septum unit : This unit computes its activation from the difference of activation between CA1 and CA3, giving this way a measure of the difference between what is recalled and what is currently perceived. It may inhibit the recurrent connections of CA3 and also the connections between CA3 and DG (see [3]).

Because of the conjonctive activity computation between EC and DG units, the DG activity pattern is more sparsed and less overlapped than the EC one [6]. The disjonctive activity computation within CA1 and CA3 will tend to preserve some of this non-overlapping property. The recurrent network present within CA3 and the hebbian learning will allow the memorization of the pattern. When septum is activated enough because the mismatch between CA1 and CA3 is too high, recurrent connections and connections from DG are inhibited, and CA3 units receive excitation from their corresponding unit in CA1, hence forcing the original pattern into CA3 to be learned again.

VI. Model of associative cortex

A. Events sequencing in the model

Having hypothetized the world was "frozen", each event that is perceived by the primary visual¿areas and memorized by the hippocampus can be linked to the next event, using the effector action as an index.

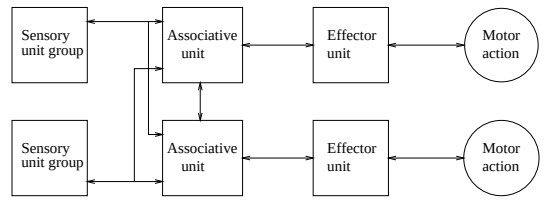

Fig. 6. The associative units are tuned to a specific effector unit which can be of any kind. These units are also linked to several sensory unit groups. There is no explicit semantic attached to these associative units, nonetheless, it is easy to find their implicit meaning considering the corresponding effector.

The model of associative cortex is composed of several units with different roles :

- Effector units : These units are hardlinked to an effective motor action in the framework of the task described earlier : up, down, left or right. Each motor action will then excite these effectors, and each excitation will induce a motor action.
- Sensory unit groups : Each group is linked to a group in the entorhinal cortex area. Typically, in the navigation framework, they are tuned to a specific set of retina perception and are then equivalent to place cells. They are also linked to each associative unit at the thalamic and cortical level.
- Associative units : These units are linked to a single effector unit and to every sensory unit groups and associative units. They are thus able link places and motor actions.

Considering the case when the robot is moved from *locus A* to *locus B* with the *up move*, the effector unit corresponding to the up move and the corresponding associative unit are then excited. The sensory unit group corresponding to locus B is higly excited whenas the sensory unit group corresponding to locus A is less excited. The associative and the sensory unit group will then reinforce their reciprocal links up to the point where the associative unit will specialize itself onto this sequence of event and will then divide itself. The newly created column will then inherit part of the links of the mother and inhibit these very links in the future, i.e. the mother column will not be able any more to detect this specific sequence.

B. Using learned sequences

Once the learning phase is over, there exist several associative units (result of splitting) linking two neighbour places. If a call is initiated somewhere on the map, this call will be propagated to all neighbour units, thanks to columnar inner mechanism. Of course, this call is not strong enough to directly excite an associative unit, nonetheless, the spread of call will soon reach the group of units which represent the current location. Then, an associative unit will be connected to both an excited unit group and to a calling unit group. It will be sufficient for it to become excited, inducing this way the corresponding motor action. Finaly, step by step, the motor sequences will lead the simulated robot to the right "calling place".

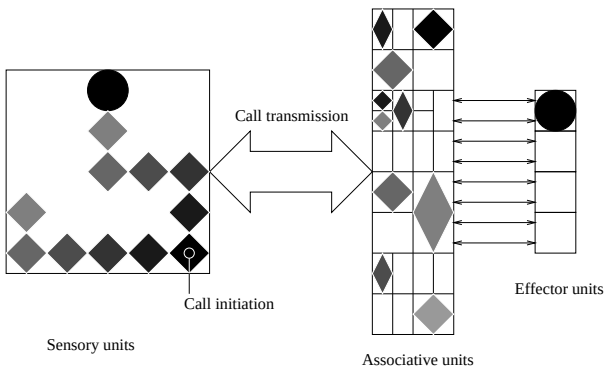

Fig. 7. Once the map is learned, a call is initiated somewhere on the map, the call propagation via the dynamic structure of the associative units will allow the call to be propagated up to the excited sensory unit group, allowing this way the action to be carried out.

VII. Perspectives

Our perspectives are to implement the model onto a real robot (Koala) in order to test its robustness. It requires both the use of a more complex retina (currently under development) which will allow a more global perception of the world and also a more accurate set of motor actions able to deal with natural noise of real environment. First, the robot will have to use only declarative memory (the cognitive map) to be able to navigate throughout the environment in order to find food and water when needed. Second, the robot will have to use both declarative and procedural memory and decide which action has to be performed. For example, if food is required and is currently perceived via the retina, the robot has to decide whereas it tries to go where food is currently perceived or if it goes where it is known there is food.

References

[1] D. H. Ballard. Cortical connections and parallel processing: Structure and function. *The Behavioral and Brain Sciences*, 9:67–129, 1986.

[2] Y. Burnod. *An adaptive neural network the cerebral cortex.* Masson, 1989.

[3] M. E. Hasselmo and B. P. Wyble. *Hippocampus*, volume 6, chapter 6, Encoding and Retrieval of Episodic Memories : Role of Cholinergic and GABAergic Modulation in the Hippocampus, pages 675–684. David G. Amaral and Menno P. Witter, 1996.

[4] D. Marr. Simple memory: A theory of archicortex. In *The Philosophical Transactions of the Royal Society of London*, number 262 in B, pages 23–81, 1971.

[5] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionnists models of learning and memory. Technical report, Carnegie Mellon University and The University of Arizona, March 1994.

[6] B. L. McNaughton and L. Nadel. *Neuroscience and connectionnist theory*, chapter 1 : Hebb-Marr networks and the neurobiological representation of action in space, pages 1–63. M. Gluck and Rumelheart, Hillsdale, erlbaum edition, 1990.

[7] P. Milner. A cell assembly theory of hippocampal amnesia. *Neuropsychologia*, pages 23–30, 1989.

[8] J. Murre. *Hippocampus*, volume 6, chapter 6, TraceLink : A model of Amnesia and Consolidation of Memory, pages 675–684. David G. Amaral and Menno P. Witter, 1996.

[9] J. O'Keefe. Place units in the hippocampus of the freely moving rats. *Experimental Neurology*, 51:78–109, 1976.

[10] J. O'Keefe and J. Dostrovsky. The hippocampus as a spatial map : preliminary evidence from unit activity in the freely moving rat. *Brain Res*, 34:171–175, 1971.

[11] J. O'Keefe and L. Nadel. *The hippocampus as a cognitive map.* Oxford University Press, 1978.

[12] E. T. Rolls. *Hippocampus*, volume 6, chapter 6, A Theory of Hippocampal Function in Memory, pages 601–620. David G. Amaral and Menno P. Witter, 1996.

[13] W. B. Scoville and B. Milner. Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, pages 11–21, 1957.

[14] L. R. Squire. Memory and the hippocampus: A synthesis from findings with rats, monkeys and humans. *Psychological Review*, 99(2):195–231, 1992.