From Perception-Action loops to imitation processes: A bottom-up approach of learning by imitation

P. Gaussier, S. Moga, J.P. Banquet (*), M. Quoy

Neurocybernetic team, ETIS ENSEA-UCP, 6 Av du Ponceau, 95014 Cergy, France

(*) Neurosciences et modelisations, Jussieu, Paris.

e-mail: gaussier or moga@ensea.fr

Abstract

This paper proposes a neural architecture for a robot to learn how to imitate a sequence of movements performed by another robot or by a human. The main idea is that the imitation process does not need to be given to the system but can emerge from a mis-interpretation of the perceived situation at the level of a simple sensori-motor system. We discuss the central position of imitation processes for the understanding of our high level cognitive habilities linked to self-recognition and to the recognition of the other as something similar to me. Another interesting aspect of this work is that the neural network used for sequences learning is directly inspired from a brain structure called the hippocampus and mainly involved in our memorization capabilities (Banquet & Gaussier 1997).

Introduction

Nowadays, robots are mainly directly programmed to solve particular tasks. In our laboratory, we are working on how a robot can program itself according to a little number of internal drives. Our neural network architecture (PerAc: Perception-Action (Gaussier & Zrehen 1995)), inspired from neurobiology, allows shape categorization and learning of sensory-motor associations. Using this architecture, a robot can learn from a teacher to isolate a particular "object" in a visual scene and to associate it with a motor behavior: reaching, avoiding by the left or the right...(conditioning mechanism (Joulain, Gaussier, & Revel 1997)). In the same way, for navigation in an open environment, we have shown that a robot can learn how to reach any position with high precision just by using the landmarks it finds in the visual scene (Gaussier et al. 1996). However, "discovery" of interesting locations or "discovery" of the correct set of sensory-motor associations for a particular task is an NP-complete problem. The learning time becomes quickly too large when the size of the problem increases.

Therefore imitation of already learned behaviors or subparts of a behavior not completely discovered is certainly the only way to allow a population of robots to learn and to find solutions by themselves. Imitation is also a good starting point to allow human-robot interactions. The autonomy of the robot requires that there is no intrusion in the robot brain during learning. The robot must be able to choose what to store. Hence, the problem becomes how to initiate the communication mechanism between the robot and the teacher. In this paper, after a brief summary of our previous works on robot population for clustering tasks, we propose a neural architecture for visual imitation and we show how to use it to teach the robot to perform a particular sequence of movements (to round in circles or to make a 8 trajectory...). By contrast to other works in this area (Berthouze, Bakker, & Kuniyoshi 1996; Matarić 1995; Hayes & Demiris 1994), our main concern is to allow an on-line and unsupervised learning. The long term goal is to realize a neural network controller inspired by biological and psychological findings about humans and animal development. We hope to validate those models and/or to make counter propositions.

Collective behavior

Several years ago (Gaussier & Zrehen 1994), we have realized a robotic experiment inspired by Deneubourg's work on sorting and clustering tasks performed by ants (Denebourg et al. 1990). As in Deneubourg's model, our robots did not take into account the interactions between the agents. All the robots acted as if they were alone. Nevertheless, because in our application building a cluster of pieces of wood created obstacles, the environment was divided in areas separating the robots. The robots then became specialized in the clustering of a particular area. The most interesting point in this work is certainly not the behavior group since with more than three or four robots the achievement of the task was longer than with two robots! These robots did not have a program that explicitly allowed

them to build large clusters. Indeed, the only robot instructions were to avoid obstacles like Braittenberg vehicles, to take an object if it was not already moving another object and to leave an object above the other object in the opposite case. With this kind of instructions the robot can build stacks of 2 objects but there is no reason to build stacks with more than 2 objects. Indeed there is the same probability to take an object from a stack or to put it in another stack. In average the size of the stacks would not change. However in the real experiment stacks are created which implies a bias in the probabilities of taking and setting down an object. That bias is linked to a perception problem. To separate obstacles from objects, a decision on the number of Infra Red sensors saturated on our Khepera robots was used. The consequence of putting an object nearby another object is that they appear as an obstacle and not as two objects. Their significance for the robot has changed. The intrusion of this Gestaltist effect (shape theory - the whole is bigger than the sum of its parts) has questioned us about how robots can learn by themselves to take into account those emergent properties of the robot/environment interactions and how to use them in learning by imitation.
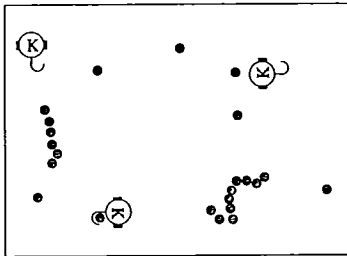

Figure 1: Overview of 'ant-like' clustering and sorting experiment.

Frame of our imitation process

For the imitation behavior, we start with the same kind of assumption than for the clustering problem: imitation is triggered by a perception error. For instance, an imitation behavior between two robotic arms controlled by vision could be explained as follows: A robot arm learns the visuo-motor coordination between its camera and its hand. It creates a correspondence between a given hand position in the visual scene and the angular positions of the different joins. Then, if that robot looks somewhere else and sees in its visual field another arm, it will perform the same movement as the second arm because it will try to reduce the differences between the representations it supposes to have of its arm (visual and motor representations). Finally, if the

arm movements allows to reduce an internal drive (associated to the refueling of the energy for instance), a positive reinforcement is triggered. The movement sequence will be stored and associated to the internal drive. Later, if the value associated to the internal drive changes too much from its optimal value, the sequence of movements will be triggered.

Robot following mechanism

Now, if we return to our mobile robot imitation problem, it is complex to imagine a program that allows a robot to learn to visualize what another robot is doing [1]. But, it is simple to allow the robot just to follow another robot as a way to avoid a difference between perception and action (reduce the difference of speed between the information of the visual flow and information about the motor wheel speed - homeostasy principle).
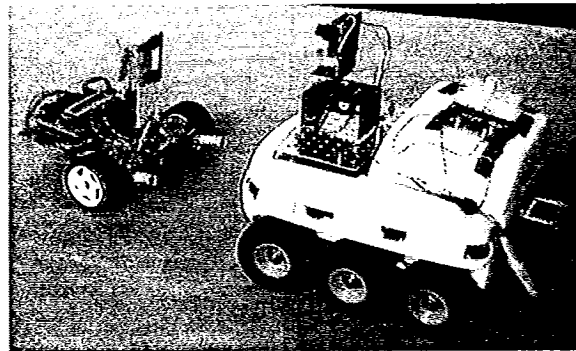

Figure 2: Overview of following sequence. On the left, the teacher robot (a MIT handy-board base), on the right, the learner robot (KOALA)

During a first learning phase, the robot could learn to go backwards when it perceives an expansion of the optical flow and to go forward in the opposite case (contraction of the optical flow). In the same way, when an expansion central point appears on the left, the robot must learn to turn right. This behavior can be learned and frozen and afterwards considered as a reflex action mechanism for the learning of more complex tasks. In our system, it is hardwired.

The extraction of optical flow is performed by the visual block (VI) of our architecture. We compute a image sequence, using a "retinal-like" model, by temporal integration of few (5 in our experiment) images from the CCD camera.

---

[1] It is something that primates and perhaps other mammals succeed to do, but in a first time, we would like to see what kind of imitation mechanism could be performed on a robot that cannot have a complex internal representation of the world.
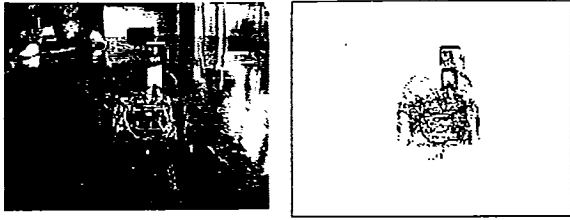
Figure 3: a) An image used by our robot in its following mechanism. b) the diffence between two time integrated images gives informations about where is the moving object.
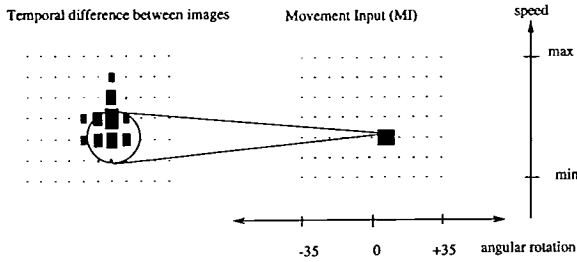

Figure 4: The result of the movement image is subsampled and the center of the maximum activity area is used to control the robot movement. MI is the Movement Input group.

A "movement image" is directly computed by thresholding the difference between 2 time integrated images of the above sequence (figure 3). The information for the robot following behavior are directly extracted from this image by keeping the same kind of egocentric coordinates. The movement direction is the value on the x axis of the neuron centered on the most activated area of movement. The speed is deduced according to the projection of the same neuron on the y axes. If the movement is just near the robot (in the lower part of the movement image) then the robot speed will be negative and the robot will try to avoid the collision as shown fig. 5. This pre-cabled mechanism allows to maintain the same distance between both robots. In our PerAc architecture (see fig. 6, it is represented by an unconditional connection between two neural maps representing the proposal of robot movement (Movement Input - MI) and the effective robot movement, a Winner Take All (WTA) group of neurons called Motor Output (MO).

Learning of a temporal sequence

Our neural network for motor sequence learning is inspired by the study of two brain structures involved in memory and time learning: the cerebellum and the hippocampus (the hippocampus is involved in the short term to medium term storage of procedural events
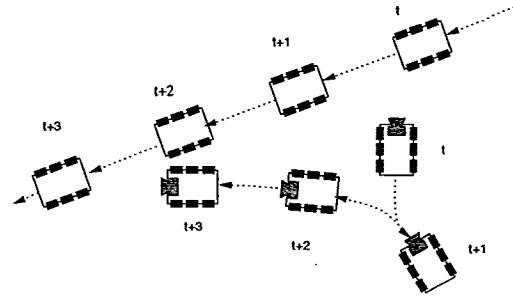

Figure 5: The reflex mechanism allows at the same time to follow a moving object or to avoid collision according to the type of optical flow (dilatation/contraction) and to apparent distance of the moving target.

(Banquet & Gaussier 1997) and the cerebellum learns motor skills like ballistic trajectories (Bullock, Fiala, & Grossberg 1994)).

Our robot does not learn directly to imitate the other robot (a home-made simple robot - see fig. 2). It just learns to reproduce its own sequences of actions primarily induced by the follow-up reflex behavior. As a result, it also learns to predict its own next movement and can use that information to detect novelty (Denham & Boitano 1996) (situations in which its predictions are wrong).
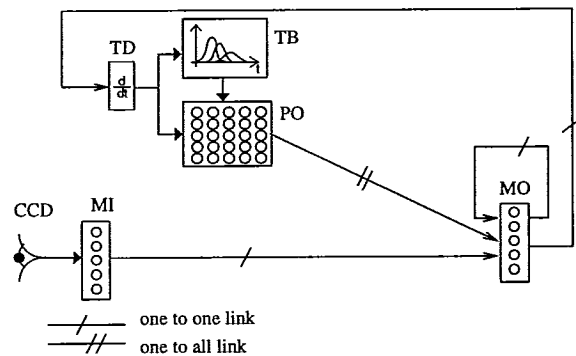

Figure 6: Overview of the PerAc architecture for robot following and sequence of movement learning.
CCD - CCD camera, MI - Movement Input, MO - Motor Output, TD - Time Derivator, TB - time battery, PO - Prediction Output

The movement changes characterized by OFF-ON transitions (Time Derivative TD group) of MO neurons are used as input information for a bank of time spectrum battery-cells (Bullock, Fiala, & Grossberg 1994) (TB in figure 6). Time battery (TB) of granular cells act as delay neurons endowed with different time constants. They also perform a spectral decomposition of the signal that will allow the neurons in the Prediction Output group (PO) to register transition patterns

between two events in the sequence. An input to a specific battery of TB granules performs both a reset of any eventual residual activity in this battery, and an initialization of the spectral timing activity within the group of cells of the battery.
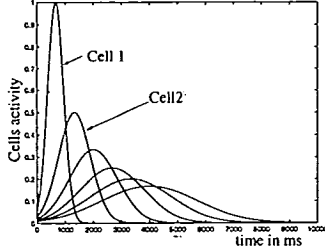

Figure 7: Time activity of a groups of cells that allows to measure time in TB.

The size of the TB battery in our application is 15. Time activity of 5 batteries of cells is presented in figure 7. The activation law of a TB cell is presented in equation 1.
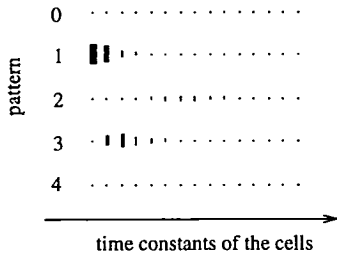

Figure 8: Time activity of TB after a sequence 2, 3, 1.

$$Act_{j,l}^{TB}(t) = \frac{1}{m_j} \cdot \exp{-\frac{(t - m_j)^2}{2 \cdot \sigma_j}} \qquad (1)$$

where j is associated to the choise of a particular time constant and l to the label of the recognized input pattern. m_j is the value of the time constant and σ_j its associated standart deviation.

The PO group receiving information from TD through both by the direct and indirect pathways (fig. 6) learns the intervals between two events of the sequence in the strength of the connections between TB and PB (proximal dendrites). More importantly, PO also encodes transition patterns between successive or separated events in the sequence.

In the proposed model, the direct input provides first a temporal information necessary to learn the temporal interval at the proximal synapses of the PO neurons. Second, it provides spatial information which help to encode the pattern of the registered activity at PO. After learning, the links between TB and PO (triggered

by the presence of a direct input from TD), the summation of the associated spectral activities provides a maximum of activity that predicts the time and type of the next event.

$$Pot_{i,j}^{PO} = \sum_{i,j,l} W_{po(i,j)}^{tb(j,l)} \cdot Act_{j,l}^{TB} + W_{po(i,j)}^{td(i)} \cdot Act_j^{TD} \qquad (2)$$

The potential is calculated as in eq. 2 and its variation trough time is presented in fig. 9.

$$Act_{i,j}^{PO} = f_{PO}\left(Pot_{i,j}^{PO}\right) \qquad (3)$$

$$f_{PO}(Pot) = \begin{cases} 1 & if \frac{dPot}{dt} < 0 \ and \ Pot > 0 \\ 0 & otherwise \end{cases} \qquad (4)$$

$$W_{po(i,j)}^{tb(j,l)} = \begin{cases} \frac{Act_{j,l}^{TB}}{\sum_{j,l}\left(Act_{j,l}^{TB}\right)^2} & if \ Act_j^{TD} \neq 0 \\ unmodified & otherwise \end{cases} \qquad (5)$$

The activation of the body of the neuron is done by equations 3 and 4. Figure 9 represents the firing of a PO neuron.
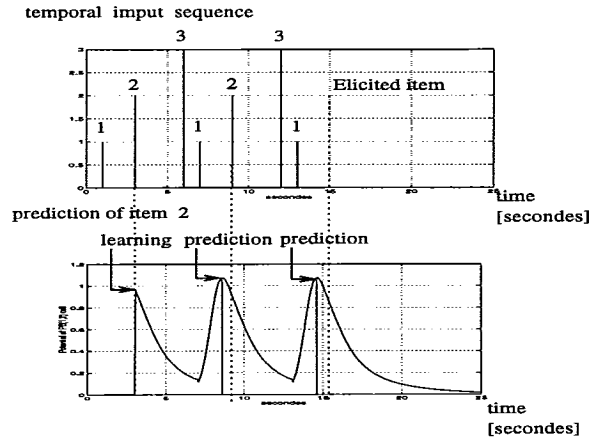

Figure 9: The potential variation for a PO cell through time. Symbols 1, 2, 3 represent the apparition time of these events.

At last, the PO group is linked to the MO group via a one-to-all link. Hence, after a first sequence of actions, a motor neuron will be activated by the reflex input and will also receive information from the transition prediction group (PO). A simple conditioning rule then allows the activated neuron to react the next time the action is predicted even if the reflex does not provide information. Moreover, performing that action

52

provides information to the event prediction mechanism that will reinforce the sequence. Thus replaying the sequence allows it to be maintained in memory. The learning mechanism needs the presentation of only one or two complete sequences of movements to learn to predict the change of movement (the system learns the timing) and to reply the correct sequence. If the sequence of movements induces a positive reward then the past predicted transitions and their associated movements are reinforced. This way, the robot learns to imitate the behavior of the other robot. It succeeds in reproducing learned sequences of movements according to the activated motivation. The detailed connectivity of the prediction part of the network is presented in fig. 10.
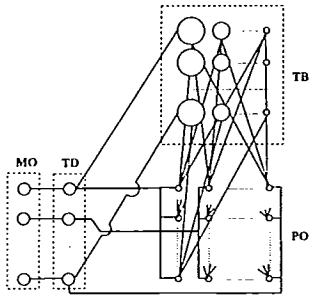

Figure 10: Detailed connectivity of the event prediction network. The circle size in TB is associated to the tme constants of the neurons.

Dynamic of the sequence learning

We now describe step by step the activation and learning of the different components of the system when the teacher performs a square trajectory.

For the sake of clarity, suppose that TD receives the sequence of patterns 1, 2, 1, 2, ... at variable intervals that could correspond to "go ahead" and "turn right" (to make a square for instance). While pattern 1 of activation is maintained in the MO register, the TD corresponding unit performs a time derivative of this pattern, and simultaneously forwards to TB and PO downstream systems an activation pattern.

There is no significant learning related to pattern 1 direct input to PO, due to the absence of any significant previous pattern through the indirect TB pathway. Conversely, the TD input to TB resets any residual activity in the selected battery of TB neurons. At the same time it triggers an activation of the different components of the battery. This activation unfolds at different paces for the different components of the battery. Thus this battery will keep a count of time until the first reoccurrence of pattern 1.

This way, a subset of the PO neurons, represented in fig. 10 by a row of the matrix, reflects the TB temporal pattern 1 of activation in a subthreshold, latent activation of PO proximal dendrites. Nevertheless, this sole TB input is not enough to trigger by itself learning at the PO level. Learning will only result from the conjunction of this TB input with a phasic transient signal from TD. In our case this direct input will now be pattern 2 of the sequence. Indeed, there is no significant learning related to pattern 1 direct input due to the absence of any significant previous pattern.

After the end of the first presentation of the sequence, when event 1 is presented again, besides the recognition of an event repetition, different transitions will be learned. Most importantly the transition from 2 to 1, but also from 1 to 1 (an event repetition). At this point of the repetition of the sequence the PO predictive capacity will become operational. As the presentation of events 1 and 2 permitted the latent learning of the 1-2 transition, similarly, the presentation of a new event "3" could provide for the learning of the 2-3 transition, and also the 1-3 transition.
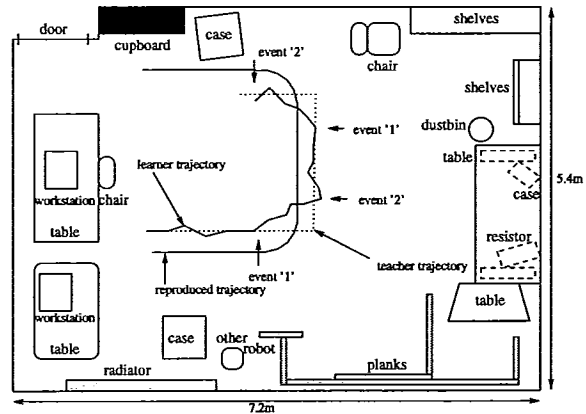

Figure 11: experimental trajectories

In the robotic experiment, the same behavior is retrieved. In fact the main problem is to filter the movement "noise" linked to the variability of the movement perception and to the limitations of the possible movements. "Noise" is also due to the fact that the learner tries to cut the trajectory of the teacher and does not provide a clear transition from one orientation to the other (the set of successive rotations to follow the other robot is not repetitible). Hence, there is a need to integrate throw time the orientation of the teacher and to consider there is an orientation change only if the variation of the integrated absolute orientation of the robot is high enough. In the robotic experiment, we have added a magnetic compass to measure the real

robot orientation (a measure of the orientation by the robot odometry can also be used) and all the learned information come from that value (values integrated through time). It allows the robot to perform a 90 degree rotation at approximatly the correct time whereas during learning it had begun to turn slowly (10 or 20 degrees per step) a long time before the good time (because the teacher was beginning to turn). The time needed to integrate the orientation information allows to learn to predict an instant very accuratelyt. (see fig 11).

Conclusion

This work is just at its very beginning. Tools to allow learning by imitation seem to be able to solve a wide variety of problems. Nevertheless, a lot of fundamental problems must be further investigated. First, how to decide the moving object is "like" the robot and so must be imitated? In our approach, everything could be imitated and after a while, if no reinforcement signal is received, the robot should learn the object is not interesting and must not be imitated (because its visual shape will be associated to an avoidance of the imitation process). The presence of more than a single moving object in the learner's visual field also supposes the existence of an attentional mechanism able to discriminate the visual moving objects in order to determine who is the teacher.

This approach could explain how a robot can learn to recognize somebody as its sort (congener) (Dautenhahn 1995) and perhaps we will be able to generalize that to the learning of the consciousness of the robot itself (Dennett 1991). This is my arm because I can predict what it will do (reward). My internal schemes involving my arm remain stable.

In our case, the robot does not know it is imitating. Do we need to realize robots that really understand they are imitating? What does it mean as modification in the proposed robot architecture? Learning social relationship needs to add the possibility of learning "together" and not only one from an other. Techniques to manage this kind of problems seems to be available but excepted few experimentations (Dautenhahn 1995), real size applications managed in a bottom-up approach have to be imagined.

References

Berthouze, L., Bakker, P., and Kuniyoshi, Y. 1996. Learning of oculo-motor control: a prelude to robotic imitation. IROS, Osaka, Japan.

Bullock, D., Fiala, J., and Grossberg, S. 1994. A neural model of time reponse learning in the cerebellum. Neural Networks 7(6/7):1101–1114.

Dautenhahn, K. 1995. Getting to know each other - artificial social intelligence for autonomous robots. Robotics and Autonomous System 16(2-4):333–356.

Denebourg, J., Goss, S., Franks, N., Sendova-Franks, A., Detrain, C., and Chrétien, L. 1990. The dynamics of collective sorting: Robot-like ants and ant-like robots. In Meyer, J., and Wilson, S., eds., Conference on Simulation of Adaptive Behavior. Cambridge: MIT Press.

Denham, M., and Boitano, J. 1996. A model of the interaction between prefrontal cortex, septum and the hippocampal system in the learning and recal of goal-directed sensory-motor behaviours. Technical Report NRG-96-01, University of Plymouth - School of computing.

Dennett, D. 1991. Consciousness Explained. Boston, Massachusetts: Brown.

Banquet, J., Gaussier, P., and J.C. Dreher, C. Joulain, and A. Revel 1997. Cognitive Science Perspectives on Personality and Emotion Chapter Space-Time, Order and Hierarchy in Fronto-Hippocampal System: A Neural Basis of Personality newblock Elsevier Science BV Amsterdam

Gaussier, P., and Zrehen, S. 1994. Avoiding the world model trap: An acting robot does not need to be so smart! Journal of Robotics and Computer-Integrated Manufacturing 11(4):279–286.

Gaussier, P., and Zrehen, S. 1995. Perac: A neural architecture to control artificial animals. Robotics and Autonomous Systems 16(2-4):291–320.

Gaussier, P.; Joulain, C.; Revel, A.; and Banquet, J. 1996. Are shaping techniques the correct answer for the control of an autonomous robot ? In UKACC International Conference on Control'96, 1248–1253. University of Exeter: IEEE.

Hayes, G., and Demiris, J. 1994. A robot controller using learning by imitation. In Proceedings of the 2nd International Symposium on Intelligent Robotic Systems, 198–204.

Joulain, C., Gaussier, P., and Revel, A. 1997. Learning to build visual categories from perception-action associations. In accepted to IROS'97.

Mataríc, M. 1995. Issues and approaches in the design of collective autonomous agents. Robotics and Autonomous System 16(2-4):321–331.

Note: The link between our N.N. and an hippocampal model can be easily obtained if one considers that TD is associated to the enthorinal cortex, TB to the Dentate Gyrus and PO to the CA3 region (for more details see (Banquet & Gaussier 1997)).