

Simulations of Dynamical Interactions for Social Learning

Pierre Andry and Philippe Gaussier* and Jacqueline Nadel**

**Neuro-cybernetic team, ETIS Lab, UPRES A 8051 UCP 6 avenue du Ponceau, 95014 Cergy, France, {andry, gaussier}@ensea.fr*

***Equipe Développement et Psychopathologie UMR CNRS 7593, Hopital de la salpêtrière, 45 Bd de l'hopital, 75013 Paris, France, Jnadel@ext.jussieu.fr*

Abstract. Often viewed as a tool for learning, imitation also has a communication purpose. In this paper we consider the interactional side of imitation, and especially its dynamic. We study simulations of interactions between two agents. We show, how improvements of an architecture designed for learning by imitation, permits to have a stable interaction: a synchronization of both agents. We also show that the dynamic of the interactions provides useful informations to build a reinforcement signal. That can be used to learn an arbitrary set of perception-action associations

1 Introduction

Imitation can be viewed as a powerful learning paradigm for real and simulated agents. As a capacity of learning by observation, imitation can improve and speed up the learning and the exploration of the sensory-motor space. For example, a robot with a mechanical arm can quickly learn from a model the relevant actions to resolve a given task (moving an object to a particular place, recharging). In our purpose of designing a neural architecture able of learning by imitation, we are interested by the first levels of imitation. We are concerned with the simplest functionality that an architecture needs to perform low-level imitations, i.e reproduction of meaningless and simple movements, without any complex notions such as intentionality, consciousness of self or others, etc.. This “proto imitation” level is very important because it permits to understand the basis of the perception-action mechanism necessary to perform more high level behaviors (Gaussier et al., 1998). But imitation also permits to investigate the social and communicational relationships between agents (Dautenhahn, 1995; Billard and Hayes, 1997; Billard et al., 1998) . As observed among very young children (Fig 1, and (Nadel, 2000)), imitation is also a tool

for gestural interactions, via demonstrations and synchronous reproductions of motor sequences. The fact that imitation is present during the early stages of learning and communications, leads us to study how our model of proto imitations could also be used to study simple and simultaneous motors interactions in the human-machine or machine-machine loop. How can the interaction benefits to the learning process? We think that when agent interact, the overall dynamic of the interactions can bring useful informations, favoring exchange of signals, stimulations or learning. Because these interactions and exchanges are common to most living entities, it could be useful to characterize some important states of the dynamic, such as attractors. In this paper, we briefly present the proto imitation mechanism as the basis of a control architecture, used to learn sequences of movements with a mobile robot. We then investigate a machine-machine interaction, where two of our simulated architectures are interconnected, producing the same sequence. We show how synchronization of both systems can be obtained from an internal modulations of their perception/action loop. The importance of synchronization as the resulting dynamic is then discussed. Finally we explain how the interaction dynamic can also bring non-explicit but important informations,

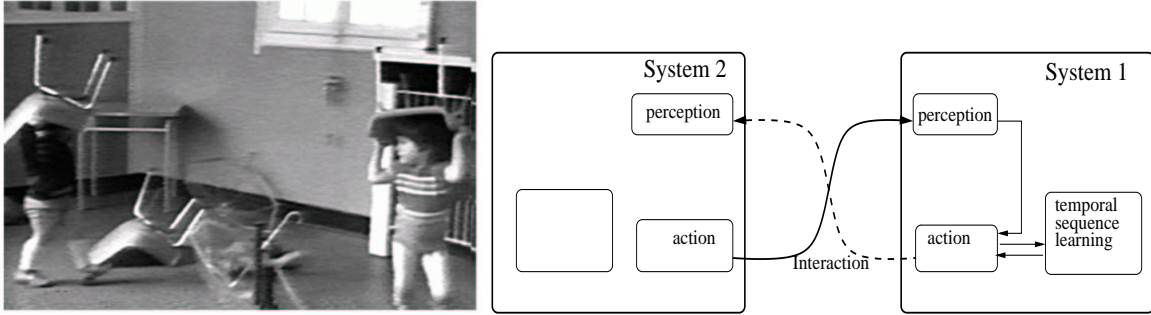


Figure 1: Young children interacting in an imitation game. Imitation deserves a communicational purpose, where synchronization and rhythm matter. Close to this idea, we study the interconnection of two systems. System 1 and 2 have the same architecture. Each system has learned associations between its inputs and outputs. The two systems produce outputs (the same sequence of motor outputs for example) at the same time.

such as the rhythm of the exchange to build an implicit reward.

1.1 Proto imitation: the ambiguity in the Perception-Action loop

We start with the assumption that imitation can be triggered by a perception ambiguity (Gaussier et al., 1998), inherent to embodied systems in a real environment. “Perception ambiguity” is a difficulty to discriminate objects (is this my arm or another’s one?), or to decide between different interpretations (is this a useful object, or an obstacle?). It was first introduced by Gestaltists, assuming that local features in a perceived scene were always ambiguous (only the global contextual information and the dynamic of the perception-action loop allow to suppress ambiguity). We think that perception ambiguity can trigger a very simple imitation behavior, such as the reproduction of meaningless movements. To illustrate our proposal, let’s consider a robot with a mechanical arm and a CCD vision system (Fig. 2). A controller of this system creates a correspondence between a given position of its hand in the visual scene and the angular position of the different joints. Let us now suppose that the robot looks somewhere else and perceives another arm in its visual field (due to a narrow field of view), and that this perception is ambiguous enough to be confused with self arm. If the observed arm moves, the controller which stored the visuo motor correspondence will then detect an error between the representations it supposes to have of its arm. Reducing this error will then make the system perform a similar movement as the observed arm. Finally if the sequence of movements is stored and associated

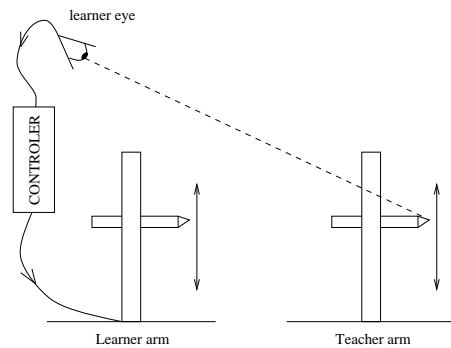


Figure 2: The student robot has already learned the correspondence between its arm internal representation and its positions in its visual field. If the student focuses its attention on the teacher arm, it will reproduce the teacher’s movement just because it will perceive a difference between the proprioceptive and the visual information. An external observer will then deduce the learner robot is imitating the teacher.

with the satisfaction of a particular motivation, it can be reproduced later. An external observer will consider that the robot has learned via imitation the behavior of the other arm. To sum up, we propose that imitation behavior can be induced by:

1. the ambiguity on the identification of the perceived arm’s extremity.
2. the minimization of the error between the visual and the motor positions(homeostasis principle (Ashby, 1960)).

Of course, the precision of the movements depends on the respective positions of the two arms. Precise

imitation would be easier if the two arms face the same direction¹. Moreover, this arm control architecture is close to the “low-level resonance” mechanism, proposed by Rizzolatti (Rizzolatti, 2000) in the “rostral part of the inferior parietal lobule (*sic*)” of monkeys. This resonance stands for the activity of the same motor neurons when observing or producing meaningless arm movements, regardless of the execution context. According to Rizzolatti, such a “low-level resonance” account for low-level imitating faculties. Perception ambiguity ensures very simple perceptions of the situation (movement and dynamics detection without any *a priori* on the shape) and allows imitation capacities without high-level notions of “self” and “others”.

1.2 Learning and reproducing temporal sequences

The idea of a link between perception ambiguity and proto-imitation is used to control a mobile robot in a kind of “dance” task. The perception only deals with detection of movements in the visual field, used to verify the rotation speed of the robot’s wheels. We obtain a sensory-motor pathway ensuring a following behavior (Gaussier et al., 1998). The robot tries to reduce the difference of speed between the information of the visual flow and the information about the motor wheel speed. As a side effect of perception ambiguity, this mechanism allows to follow moving objects. But, more than an immediate imitation, the system also learns the whole trajectory made by the teacher, in order to reproduce it later. Here, the “whole trajectory” stands for the entire temporal sequence of motor actions (see (Gaussier et al., 1998; Moga and Gaussier, 1999) for more details about the architecture, the tracking system and the filtering problems). The robot learns to reproduce its own sequence of actions primarily induced by the tracking behavior. Inspired by two brain structures, (the cerebellum and the hippocampus, see (Banquet et al., 1997) for more neuro-biological references) the network which learns the timing of the sequence of actions is also able to predict it, in order to reproduce it fidelly (Reiss and Taylor, 1991). Hence, the robot learns to imitate the behavior of the teacher (robot or human) and succeeds in reproducing the sequences ((Moga and Gaussier, 1999)).

¹such as a sport teacher showing how to perform a particular movement

2 Synchronization of actions in an imitation game

Once motor sequences are learned, they can be used, in order to solve a given problem, but they also allow the system to participate in non verbal gestural interactions. Such a system could interact with others (robots or humans) by producing, demonstrating and sharing motor sequences. This situation is inspired by young children interactions and games, where immediate gestural imitations are performed quasi-simultaneously, involving capabilities such as motor synchronization, where the same sequences of actions are executed at the same time. The synchrony can be seen as an attractor of a cyclic interaction game between two or more agents. We propose to study the overall dynamic of a loop made of two identical systems, with perception and action groups interconnected, in order to understand which minimal features need to be added to our architecture. In the following text we explain how minor improvements of the architecture (Fig. 3) can solve the trade-off between independent production and adaptive production according to the other, *i.e.* synchronization. The developed solution is inspired from the “entrainment” phenomenon observed by C. Huyggens in 1665, in which two pendulum clocks placed on the same support synchronize themselves (“clock synchronization”). Here, perception is similar to the physical wave transmitted by the support, and must allow to add some energy to the system in order to trigger the motor output earlier. Perception is adding energy to the system’s actions. Let us suppose that both architectures use the “transition learning and prediction” mechanism detailed in (Moga and Gaussier, 1999; Andry et al., 2000; Gaussier et al., 1998). Perception and Motor groups are highly simplified (binary values) since we study the networks in a computer simulation. The output of the first system is connected to the input of the second one, and vice versa (this simulates perfect perceptions of the other’s action). Both systems learn the same sequence of actions (for example transitions $1 \rightarrow 2$, $2 \rightarrow 3$, $3 \rightarrow 1$ are learned, allowing the production of the sequence $1,2,3,1,2,..$ and so on). The direct pathway connecting perception to action must be inhibited (Fig 4), in order to permit an independent and complete production of the sequence (otherwise perception and action of both system would interfere each other). The inhibition is done by connecting \overline{ND} neurons (“novelty non detection”) to the perception (Input) group. Once activated, these neurons stay potentialized (due to the recurrent

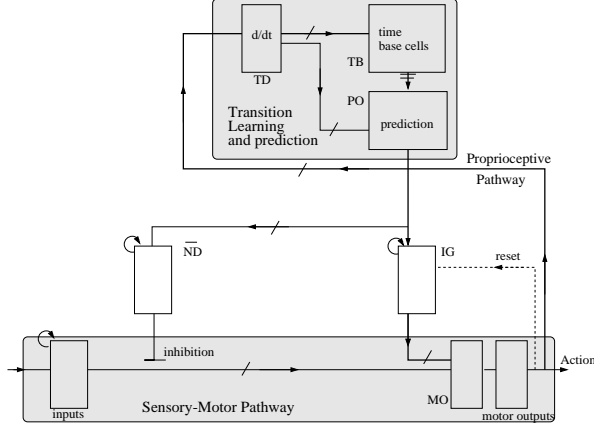


Figure 3: The model, allowing proto imitation. The system is designed as a PerAc block. The “transition learning and prediction” mechanism is a perception level modulating the “sensory-motor pathway”. Introduction of new elements allow synchronization between agents. “Non novelty detection” (\overline{ND}) and Integration (IG) groups are used to control the internal dynamic of the system.

connections) and inhibit the corresponding input’s neuron. Inhibition is triggered by Prediction (PO) neurons, which act as “decision cells” about the end of the learning phase (was this event learned or not?). Equation 1 shows the computation of the potential of the \overline{ND} cells (see the appendix, eq 8 for the computation of the activity).

$$\frac{dPot_i^{\overline{ND}}}{dt} = -\alpha_{\overline{ND}} \cdot Pot_i^{\overline{ND}} + \sum_l W_{\overline{ND}(i)}^{PO(i,l)} \cdot Act_{i,l}^{PO} \quad (1)$$

$$Act_i^{\overline{ND}}(t) = \begin{cases} 1 & \text{if } Pot_i^{\overline{ND}}(t) > 1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where α is the value of the recurrent connection. A strong inhibition of the sensory-motor pathway cuts the system from the interaction. The system produces its sequence independently, and “blindly”. The inhibition must therefore only be partial: the input signal must not be strong enough to induce motor reaction, and must nevertheless be present in order to modulate the motor production. Here, “modulate” stands for taking a relevant perception into account in order to change the timing of the sequence. Precisely, the perception of a given action, could help accelerate the triggering of the corresponding action by the system. Minor improvement of the architecture permits to obtain the modulation of the speed reproduction, in the manner of a Phase Locking Loop (PLL). The acceleration mechanism is based on the following idea: acceleration is

due to a fusion between an anticipatory information of the next motor event, and the perception of this motor event. In other words, if the system “knows” in advance the next action to perform, an incoming perception of this precise action could trigger it earlier. Two mechanisms are involved. First, a modification of the connections between time base (TB on Fig 3) and PO groups (a standard conditioning or least mean square (LMS) rule, see eq 3) permits to have an early prediction of the next transition.

$$\frac{dW_{PO(i,j)}^{TB(j,l)}}{dt} = 2\mu \cdot (Act_i^{TD} - Pot_{i,j}^{PO}) \cdot Act_{j,l}^{TB} \quad (3)$$

The early prediction is then maintained by an integration of the PO output. This is done by integration recurrent neurons (called IG, see Fig 4) according equation 4. The potential of these integrators linearly increase since prediction is emitted, and is then send to the motor (MO) group of neurons.

$$\frac{dPot_i^{IG}}{dt} = \alpha_{IG} \cdot Pot_i^{IG} + \sum_l W_{IG(i)}^{PO(i,l)} \cdot Act_{i,l}^{PO} - reset \quad (4)$$

$$reset = \sum_l W_{IG(i)}^{MO(i,l)} \cdot Act_{i,l}^{MO} \quad (5)$$

$$Act_i^{IG}(t) = Pot_i^{IG}(t) \quad (6)$$

where α is the value of the recurrent connection, H the Heaviside function and $W_{ig(i)}^{mo(i,l)}$ a strong negative weight (a motor activation resets the IG group).

The effective timing of the transitions has also to be learned (in case of an alone reproduction): a conditioning (LMS, eq 3) rule also permits this learning on IG-MO connections. Then, MO neurons simply sum 2 incoming informations (eq 7): the IG activity, which is enough to trigger activation, when it overcome a fixed threshold² and the partially inhibited perception (from Input group), which level is not enough to overshoot the threshold.

$$Pot_i^{MO} = \sum_j (W_{i,j}^{IG} \cdot Act_j^{IG} + W_{i,j}^{Input} \cdot Act_{i,j}^{Input}) \quad (7)$$

$$Act_i^{MO}(t) = \begin{cases} 1 & \text{if } Pot_i^{MO}(t) > \text{threshold} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

If a given action is perceived after its prediction, then it will increase the potential of the correspondent MO neuron, overshooting earlier the threshold: the system accelerates. Perception is then acting as an addition of energy on the system, triggering earlier the corresponding action. Between two systems

²alone, the system is able to reproduce a learned sequence

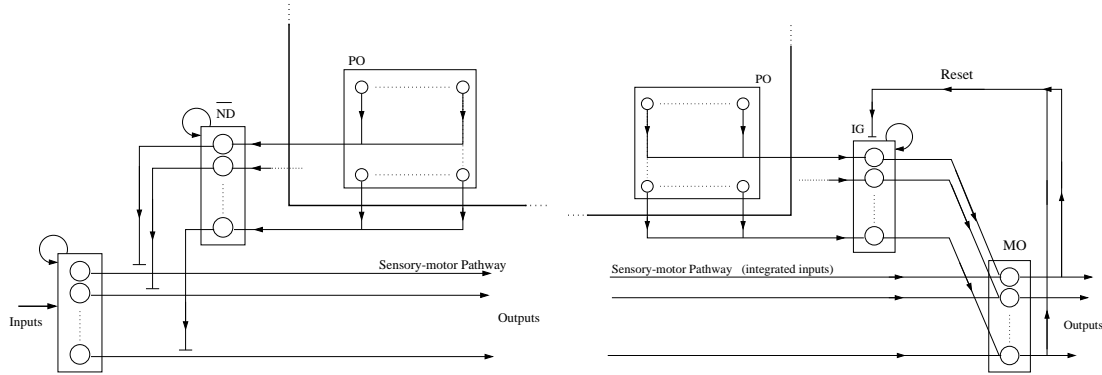


Figure 4: **Left:** \overline{ND} neurons permits to inhibit a given input. The triggering information comes from the corresponding PO neurons, firing if the corresponding event is implicated in the production of a sequence. **Right:** The Motor group (MO) triggers the motor output when integrated prediction overshoot a given threshold. If an input information (also integrated) happens during the integration of prediction, the summing potential of MO reach the threshold early: the system accelerate.

producing the same sequence, the effect of connecting action to perception induces a step by step adjustment of the sequence production until synchronization. Simulations were realized according to two experimental protocols. A first series of experiments tests the synchronization of our architecture with a given teacher. This “teacher” is a simple generator of time fixed sequences (without adaptation). Once the sequence is learned, we compare the system and the teacher’s production triggered at a random instant. In a second series of experiments, two systems are used. Both have learned the same sequence (from a fixed sequence generator). Both systems are switched on at random instants. Each architecture is a separate application, running under PVM³. For both experiments, synchronization is an attractor of the interactions. The conditions and speed of convergence toward this attractor are directly dependant of the detailed equations. Synchronization is here an example, of a stable state which can be obtained by controlling parameters and equations of the systems forming parts of the whole dynamic.

3 The dynamic of the interaction can carry useful informations

In (Andry et al., 2000), we show how the “transition learning and prediction” network can be used

³PVM, (Parallel Virtual Machine) is a software to create a virtual machine from a set of computers

to extract relevant information from the dynamic of the interaction. The application concern the learning of a set of arbitrary associations. In a simple association problem, sensory input as to be learned associated to a given motor response (inputs and outputs are coded by a single neuron). The use of reinforcement rules permits to change the values of the weights, in order for the system to learn the good associations. But what happen if the system does not have access to any explicit reinforcement or reward value? In a dynamical process of interaction (exchange of “motor” actions) between human and machine, we show that our architecture can build a non explicit reinforcement value from the rhythm of the game. The system discover by itself the reinforcement information in the data flow which is only a temporal series of inputs. The prediction mechanism (see (Andry et al., 2000) for more details and equations) can be used to build an internal reward efficient enough to control a simple reinforcement learning mechanism. The network learns the time delay between 2 keys pressed of the teacher. Once the delay is learned (after the second action of the game) the system is able of maintaining a temporal prediction of the next key pressions.

By matching the prediction information on the incoming signal, we obtain a reinforcement value, used to modulate dynamically the learning of a set of the associative network. Fig 7 shows a whole trial, with the input train of the human key activations. For each one, the system gives a response (randomly at start). Then, it learns from the teacher’s behavior. Interruptions in the human play are visible, and the middle graph shows the suc-

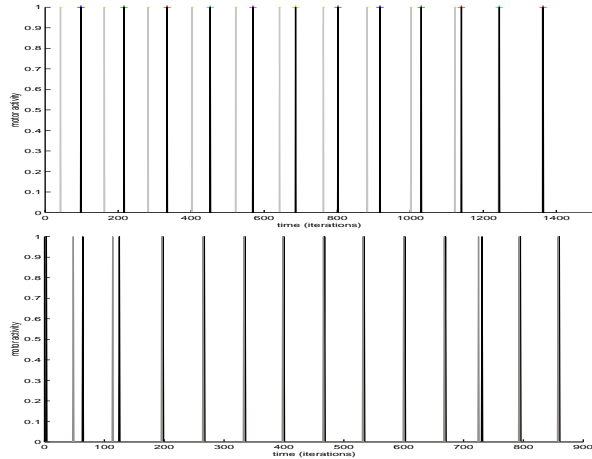


Figure 5: Interaction between two systems: Upper graph shows the synchronization phase of our system (grey) on a constant producing teacher (black). Lower graph shows the very quick synchronization between two similar systems (grey and black). A small perturbation occurred at nearly iteration 730, due to the temporary loss of messages between stations (via PVM). Perfect synchronization is quickly recovered

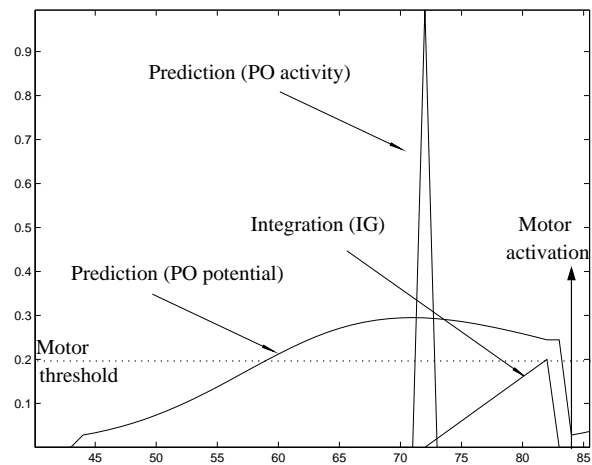


Figure 6: Activity of PO and IG cells during the production of a sequence: PO cell fires for a transition. The maximum of its potential generates an activation spike. This spike will be integrated (IG), until the overshoot of the MO neurons threshold, triggering the motor activation.

cess/error information extracted by the prediction system. Experiments lasts about 4 minutes. The network always succeeds to learn the 3 good associations. Most of the time, an experiment starts

with wrong presentations associations due to the random weights initialized on the network links, or wrong learned ones. This explains the numerous player's breaks (Fig 7, right). Here the system is capable of detecting novelty (a non expected rhythm break) in the way the teacher "plays". In further robotics implementations, we will try to extend this concept of novelty, which can be present in temporal contingencies, but also in the signal itself. This could help to "locate" important events or, to learn properties on objects or surrounding entities. This interest for novelty could motivate the reproducing of the behavior, and lead the learning process of our robots.

4 Conclusion

The imitative actions of our system are all based on an homeostatic principle. The variations of the perceptive flow are interpreted as a signal of error. In an effort to reduce the signal of error, the system induces an imitative behavior. We believe that this simple principle can be useful to build systems that are really able to process in an autonomous way different levels of imitation problems, such as following a path, reproducing the movement of an arm, or imitating complex actions like opening a door, cooking, or building a complex machinery. Another distinctive feature of the homeostatic principle is to generate predictions about future perception of action. If the system is able to predict correctly every next event, the equilibrium regime is achieved, thus suggesting the end of the learning phase. When the system has to perform the immediate imitation of an already learned behavior, uncoupling perception and action appeared to be decisive in that this uncoupling prevents perceptive interferences. Although current assumption underlying robotic works on imitation is that distinctive capacities should characterize teacher and imitator systems, our model shows the interest to implement the two systems with similar basic imitative capacities. Indeed, if the two systems come to perform a similar action, this induces a phase lock on this action, which will facilitate further introduction of turn-taking and learning from each other.

In this work, the collaboration with a psychologist has been crucial, since we have tried to reproduce some of the fundamental properties of young infant imitation capabilities such as the immediate imitation. We have discovered that immediate imitation was much more complex than previously thought and that its double role in learning and communication was of a very high importance

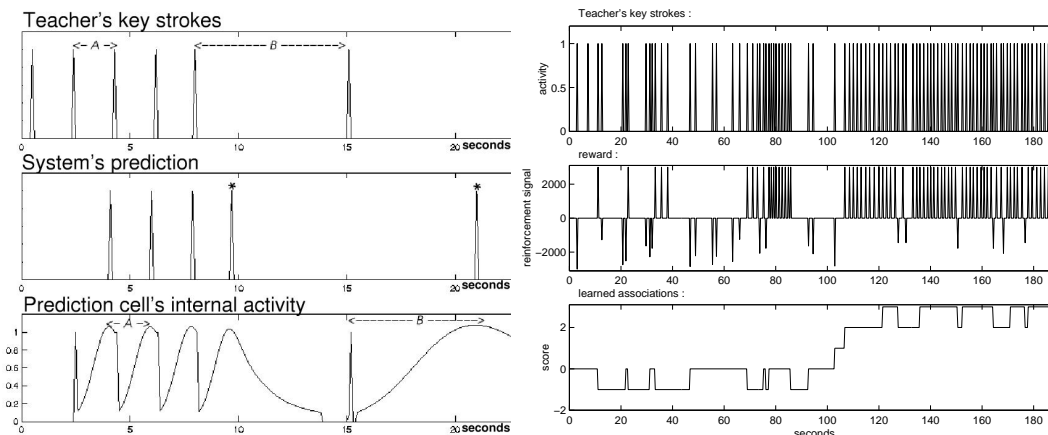


Figure 7: **Top:** Example of a 20 seconds experiment scenario. Upper graph: Each pulse indicates when the teacher press a key. A constant rhythm is maintained for the first five strikes (time period = A), then is broken. Middle graph: Pulses show the system’s binary prediction of the teacher’s play. The stars indicate a prediction error due to the break. Bottom graph: activity of the PO cell. Maximums of activity trigger pulses of prediction. After the rhythm break, the new delay, B, is learned. **Bottom:** The learning process during a whole experiment. Upper graph: action frequency of the human player. Middle graph: variation of the reinforcement signal according to the rhythm prediction. These reinforcement variations act on the update of the associative weight. Bottom graph: Evaluation of the learning level. A well learned association gives 1 point, an unlearned one gives 0 point while a bad learned one subtracts 1 point.

for robot learning. The idea that an autonomous system could generate by itself its own reward and avoid the need of an explicit reward signal also come from psychological studies. Indeed, we believe that when babies prefer to look at a screen presenting their mother interacting with them in real time instead of the delayed video of their mother, it means something very important about the kind of information babies use to detect the adequation of their behavior with their care-givers.

Finally, we are interested in the idea that imitation capabilities could be progressively build-up from very simple sensory-motor schemes, since it would promote important advances in man-machine interface and robot learning. Moreover, a robot could be considered as a good heuristic tool to propose new behavioral therapies. Modeling the conditions which are necessary to imitate is expected to suggest computerized monitoring aimed at stimulating the adequate level of imitation achievable, given the present imitative capacities in children with autism, who are supposed to face problems with human models. Our future works will focus on real size robotic experiments using mobile robots with mechanical arms. We will have to understand how to add in our neural model structures allowing to learn categories of actions at the program

level. We hope the developed architecture will be able to exhibit different phases of developments that we will be able to compare with babies development. Hence, we will perhaps be able to help in the understanding of mental development problems like children with autism: is autism linked to a theory of mind problem ? or is it linked to a more sensory-motor level or to some problem in the management of novelty detection or the capability to mobilize or express internal states ? This kind of questioning is perhaps strange for engineers since we know we are really far away from building non autistic robots. But, It is clear that all the new results in psychology and neuro-imagery will be of high interest to improve current robot controllers. At the same time, robotics experiments appear more and more as a new way to perform synthetic simulation of psychological and neurobiological models and are promised to an important development in the field of cognitive sciences.

A Notations

$W_{A(i,l)}^{B(k,m)}$ stands for the weight of the connection between the neurons of A and B groups. i, l, k, m indicates precisely the position (line and column) of

the neuron in the group.

References

- Andry, P., Moga, S., P.Gaussier, Revel, A., and Nadel, J. (2000). Imitation : learning and communication. In Meyer, J. A., Berthoz, A., Floreano, D., Roitblat, H., and Wilson, S., editors, *Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior SAB'2000*, pages 353–362, Paris. The MIT Press.
- Ashby, W. R. (1960). *Design for a brain*. London: Chapman and Hall.
- Banquet, J., Gaussier, P., Dreher, J., Joulain, C., and Revel, A. (1997). Space-time, order and hierarchy in fronto-hippocampal system : A neural basis of personality. In Matthews, G., editor, *Cognitive Science Perspectives on Personality and Emotion*, pages 123–189. Elsevier Science BV Amsterdam.
- Billard, A., Dautenhahn, K., and Hayes, G. (August 1998). Experiments on human-robot communication with robots, an imitative learning and communicating robot. *Proceedings of "Socially Situated intelligence" Workshop, part of the Fifth International Conference on Simulation of Adaptive Behavior 98, SAB 98*.
- Billard, A. and Hayes, G. (July 1997). Transmitting communication skills through imitation in autonomous robots. *Proceedings of Sixth European Workshop on Learning Robots, EWLR97*.
- Dautenhahn, K. (1995). Getting to know each other - artificial social intelligence for autonomous robots. *Robotics and Autonomous Systems*, 16(2-4):333–356.
- Gaussier, P., Moga, S., Banquet, J., and Quoy, M. (1998). From perception-action loops to imitation processes: A bottom-up approach of learning by imitation. *Applied Artificial Intelligence*, 7-8(12):701–727.
- Moga, S. and Gaussier, P. (1999). A neuronal structure for learning by imitation. In Floreano, D., Nicoud, J.-D., and Mondada, F., editors, *Lecture Notes in Artificial Intelligence - European Conference on Artificial Life ECAL99*, pages 314–318, Lausanne.
- Nadel, J. (2000). The functional use of imitation in preverbal infants and nonverbal children with autism. In A.Meltzoff and Prinz, W., editors, (*in press*). *The Imitative Mind: Development, Evolution and Brain Bases*. Cambridge: Cambridge University Press.
- O'Keefe, J. (1991). The hippocampal cognitive map and navigational strategies. In Paillard, J., editor, *Brain and Space*, pages 273–295. Oxford University Press.
- Reiss, M. and Taylor, J. G. (1991). Storing temporal sequences. *Neural networks*, 4:773–787.
- Rizzolatti, G. (2000). From mirror neurons to imitation: Facts and speculations. In A.Meltzoff and Prinz, W., editors, (*in press*). *The Imitative Mind: Development, Evolution and Brain Bases*. Cambridge: Cambridge University Press.
- Thompson, R. (1990). Neural mechanism of classical conditioning in mammals. *Phil. Trans. R. Soc. Lond. B*, 329:161–170.
- Whishaw, I. and Jarrard, L. (1996). Evidence for extrahippocampal involvement in place learning and hippocampal involvement in path integration. *Hippocampus*, 6:513–524.