# From Visuo-Motor Development to Low-level Imitation

**Pierre Andry*** **Philippe Gaussier *** **Jacqueline Nadel****
* Neurocybernetic team, ETIS Lab, UPRES A 8051 UCP-ENSEA
ENSEA, 6 avenue du Ponceau, 95014 Cergy, France
** Equipe Développement et Psychopathologie UMR CNRS 7593
Hopital de la Salpétrière, Pavillon Clérambault
54 bd de l'Hopital, 75005 Paris, France
andry@ensea.fr, gaussier@ensea.fr

## Abstract

We present the first stages of the developmental course of a robot using vision and a 5 degree of freedom robotic arm. During an exploratory behavior, the robot learns visuo-motor control of its mechanical arm. We show how a simple neural network architecture, combining elementary vision, a self-organized algorithm, and dynamical Neural Fields is able to learn and use proper associations between vision and arm movements, even if the problem is ill posed (2-D toward 3-D mapping and also mechanical redundancy between different joints). Highlighting the generic aspect of such an architecture, we show as a robotic result that it is used as a basis for simple gestural imitations of humans. Finally we show how the imitative mechanism carries on the developmental course, allowing the acquisition of more and more complex behavioral capabilities.

## 1. Introduction

In (Gaussier et al., 1998) we started to present the developmental course of an autonomous mobile robot, based on imitative capabilities. Firstly, we showed how a neural network architecture could learn by imitation sequences of actions based on a reflex sensory-motor repertory. We then showed how the new repertory of sequences of action could be used to induce simple gestural interactions with humans or other robots (Andry et al., 2001) opening new communicational perspectives. To tackle more intuitively these new issues and to increase the space of the possible associations, we equipped two of our mobile robots, with a 5 Degrees Of Freedom (DOF) robotic arm. Consequently, the initial reflex repertory (at the basis of the sequence learning) of our mobile robot becomes unsuitable for the control of a robotic arm. Thus, as a first stage of our developmental sequence, our robot needs to fill the gap between non controlled moves and basic sensory-motor coordination, in order to imitate simple gestures: the constitution of a simple but coherent and reliable visuo-motor space. Indeed, designing a robot control architecture using vision and a robotic arm, to solve the reaching of a 3-D position in a surrounding space is not a new issue (Marjanović et al., 1996). It is nevertheless a crucial and generic problem, underlying many issues, from sensory-motor coordination (Schaal et al., 2000) to object grasping and manipulation (Niemeyer and Slotine, 1988), or dynamic gestural interactions such as imitation (Cheng and Kuniyoshi, 2000, Billard, 2001). Such an architecture should allow the robot to easily generate the appropriate movement of its arm to reach a given visual point of its surrounding space, and conversely the vision could easily focus/bypass the extremity of the arm. Far from complex recognition systems, we propose a neural network architecture able to learn and store association between 2-D elementary vision and the 3-D working space of a robot's arm. We will show how the use of a new type of associative coding neurons coupled with dynamical equations successfully solves the reaching of areas of the working space (with respect to our monocular limited vision system), independently of the number of degrees of freedom of the arm (and their potential redundancy). We will also emphasize the importance of the on-line learning process, where the robot performs sensory-motor associations thru a random exploration of its own motor dynamics and physics. As a robotic

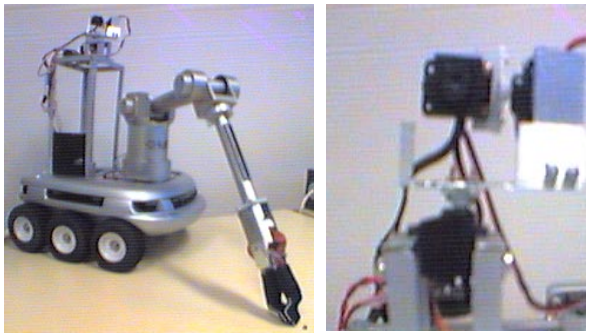

Figure 1: The robot. A Katana robotic arm and a home-made pan tilt camera (right) are mounted on a mobile Koala robot (left).

validation, we will present how dynamic and perceptive properties of such a generic architecture allow to exhibit real time imitations gestures. Finally we will discuss how the learned associations coupled with an interaction process such as imitation constitute the basis of the building of the next stages of a whole developmental course.

## 2. Developmental motivation

At birth on, vision and motor control of the neonate are relatively coarse. The young infants perceive outlines of figures, with a particular sensibility to movements in peripheral vision (Hainline, 1998). They are able to follow a moving target with saccadic eyes movements, but the motor control of the neck and the other muscles will come progressively later. This stage of development could be in a way compared to our embodied robots: they have a pan-tilt camera, a mechanical arm and a gripper, they can move, they have sensors and a CCD camera to perceive their environment (fig. 1). Nevertheless, designing a control architecture for such a robot remains complex, due to the amount of possible sensory-motor associations related to the multiples degrees of freedom. Learning to coordinate its arm to reach and grasp a visible object, to move or to push it to a given place, being able to imitate very simple gestures or movements, are all complex issues, often referring to separate works in the field of robotics. In another way, the young developing infant is able to solve all these tasks around the age of six month (Vinter, 1985, Nadel and Butterworth, 1999), and quickly uses these new skills as a new repertory for more complex tasks. More importantly, this tuning is essentially achieved during sensory-motor exploratory behaviors. These behaviors consist in repeated movements and the matching of the corresponding perceptions. These movements are not necessarily goal oriented and often randomly triggered, consisting in a first exploration of the motor space. Slowly, a coarse to fine categorization of the sensory-motor space is learned, standing for the new building blocks of a more complex sensory-motor repertory. This developmental process constitutes an interesting bottom-up guideline for the design of autonomous robots. It suggests that stable sensory-motor associations can be learned during a self-triggered exploration of the environment (here, the working space). The importance of such a developmental course would be useful for complex robots. It would allow to learn and categorize correctly the sensory motor space, according to their own embodiment, dynamics and physics.

## 3. A Neural network architecture for Visuo-Motor development

In 1989, Edelman (Edelman et al., 89) proposed DarwinIII, a simulated robot with a mobile eye and a 4DOF arm able to reinforce the reaching of particular targets. A sensor placed on the extremity of the simulated arm helped the computing of the reinforcement of the movements. Bullock et al (Bullock et al., 1993) presented the DIRECT model, a self-organizing network for eye-hand coordination learning correlations between visual, spatial and motor information. This robust solution allows successful reachings of spacial targets by a simulated multi-joint arm. Our neural control architecture is inspired from these works, and try to apply them in the context of a real robot in interactions with humans and other robots. Our Koala robots (fig. 1) are

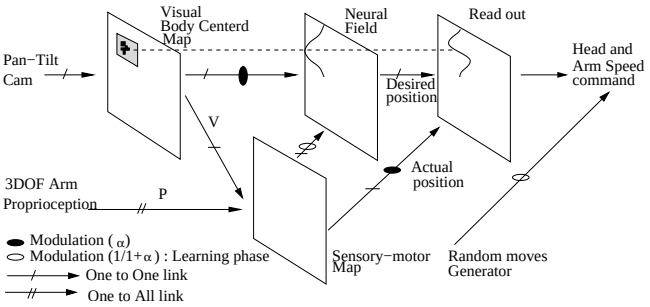


Figure 2: The simplified architecture. Proprioception of the arm and vision from the CCD camera are merged in a *sensory-motor* map composed of clusters of neurons which learns the visuo-motor associations according to visual informations ($V$). After learning, the arm proprioception ($P$) triggers the correct activity of the sensory-motor map and can be used to compute the right movement to reach a possible target using the Neural Field (NF) dynamic properties.

equipped with a 180 degrees pan and tilt monocular camera "head" (no stereo vision). The mechanical arm is a 5 DOF Katana robotic arm, but only 3 DOF will be concerned in this paper: one turret joint allowing horizontal rotations ($\theta_1$ angle), and two joints allowing verticals rotations ($\theta_2$ and $\theta_3$ angles, the two other DOF are related to the gripper control). The visuo-motor control architecture (fig. 2) is built as a perception-action loop. Two main perceptive pathways process information from vision ($V$) and arm proprioception ($P$). They are merged in a *sensory-motor* map composed of *clusters* of neurons which learns the visuo motor associations (see description in section 3.2). Visual information triggers two 1-D *Neural Field* (NF) maps whose equations compute a dynamical attractor centered on the stimulus (see description in section 4.). The direct output of the NF is then used to compute the motor command of all the devices of the robot (the head's motors and the arm's joints) to achieve, for example, pointing or tracking behaviors.

### *3.1 A simple perception system*

The conjunction of the pan and tilt motors (180x180 degrees) with the CCD field of view (35x55 degrees) allows a wide perception of the surrounding working space.

Perceptions are processed in a 2-D *camera-centered* referential. The results of this computation are then simply projected on a 2-D *body-centered* map of neurons representing the whole visual working space (fig 3). The body centered map is only a reconstruction of all the possible views that the pan tilt mechanism offers. The field of view of the camera is translated according the pan and tilt position of the head. Our robot has only a flat and two dimensional visual perception of its environment, without assumption or reconstructions of the outside world. For the experiments described in this paper, we use mainly movements detection to compute the position of the extremity of the arm. Indeed, an elementary process such as movements detection is sufficient to extract the position of the extremity of the arm for most human and robot gestures. The end point of a moving arm is naturally the most moving area, due to the summation of the angular speed of each joint. Thus, even if movement is perceived on the whole arm, in most of the cases the maxima of the movement intensity will be located on the hand. To localize the center of this area, we use a common *Winner Take All* (WTA) mechanism operating on two 1-D projections of the 2-D movement map, allowing a reliable detection of the vertical and horizontal position of extremity of the moving segment(fig 4). Moreover, we also assume that this

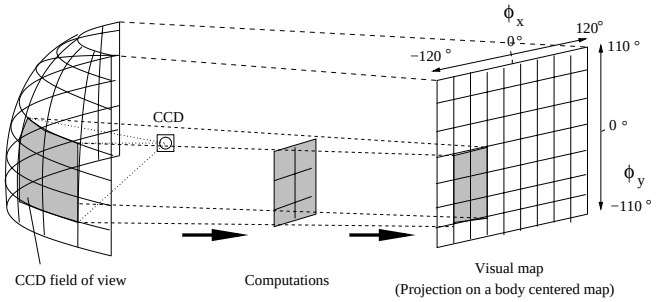


Figure 3: Mapping the CCD image on a body-centered visual map. Computations are made on a camera-centered map of neurons. The result of the computation is then projected on a body-centered map of neurons representing the whole working space of the robot

simple process is very informative since it can be performed in real time and preserve all the dynamic of the perceived stimuli [1]. The movement detection algorithm is simply a real time computation of the difference of intensity between summed packet of images, pixels by pixels (more details on the algorithm can be found in (Gaussier et al., 1998)).

[1] We have developed more complex and robust networks learning the shape of an object, but they are out of the scope of this paper

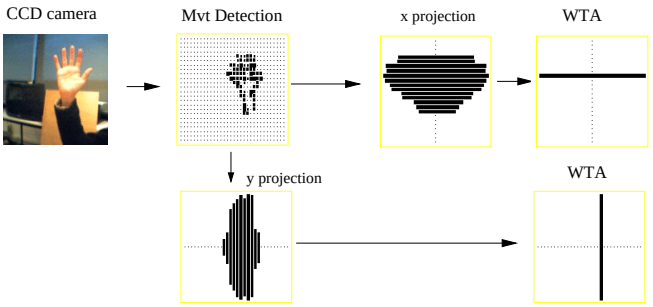


Figure 4: Example of end point tracking (here a hand) using movement detection. The movement detection (on the center) is computed from the image flow (here, the experimenter was waving its forearm). The activity of the 2-D map is projected on two 1-D maps of neurons. Then, each projection map is connected to a WTA computing the position of the maximum of movement in the scene (Computation performed at 20 images /s).

### *3.2 Learning Visuo-motor Associations*

Self-Organized Maps (SOM) such as Kohonen (Kohonen, 1976) networks are often mentioned as a possible solution for the learning of visuo-motor coordinates with simple robotic (or simulated) arms. Intuitively, the self-organizing and topology features of the network should allow a reliable learning with a reduced amount of movements during training. A drawback of this model is that it will work under the assumption that there is a bijection between the angle of the joints and the position of the extremity of the robot's arm. A given neuron activated by vision (input) can code for a given vector position of the different joints and the neighbors will learn neighbors vector position thanks to the topology of the network. But with more complex arms (such as the one we use), the same position of the extremity corresponds to multiples vector positions of the joints. To allow the association of multiples proprioceptive vectors with a single visual perception, we use a new kind of sensory-motor map, composed of *clusters* of neurons (fig 5). Each cluster of this map associates a single connection from one neuron of the visual map with multiples connections from the arm's proprioception. Visual information ($V$) controls the learning of particular pattern of proprioceptive input ($P$). Thus, this sensory-motor map has the same global topology as the visual map. More precisely, a cluster $i$ is composed of (fig 6):

- one input neuron $X_i$ linked the visual map. This neuron respond to the $V$ information and triggers learning.
- One *submap*, a small population of $Y_i^k$ neurons ($k \in [1, n]$) which learn the association between 3-D proprioceptive vectors and one 2-D vision position.

This population is a small topological map with self-organizing properties, such as a SOM maps.

- one output $Z_i$ neuron, merging information from $X_i$ and $submap_i$.

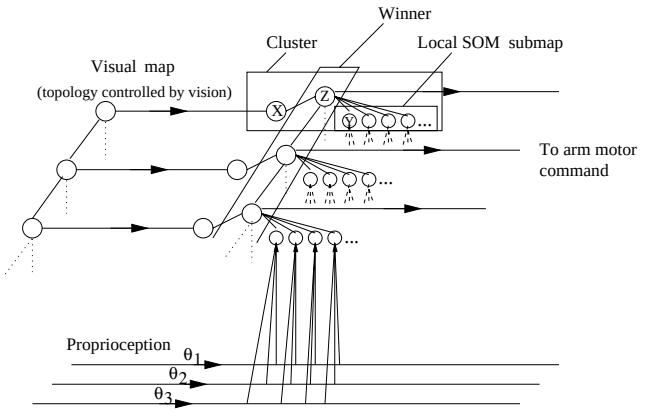


Figure 5: Organization of the sensory-motor map. The map has the same dimensions (2-D) as the body-centered visiual map. Thus the activity of one neuron of the vision map will trigger the learning of the corresponding cluster of the sensory-motor map (one to one links).

This organization ensures a topological and independent learning of each cluster stimulated by the presence of a visual input. This auto-supervised learning process allow stable and coherent learning of the visuo-motor associations. A submap of neurons is computed exactly as a simple Kohonen map. The activity of $Y_i$ neurons is proportional to the distance between $P$ values and the weights connected to proprioceptive inputs (eq 1).

$$Y_i^k = \frac{1}{1 + |P_j - W_{kj}|} \quad (1)$$

Where $Y_i^k$ is the $k$th neuron of the submap associated to the $i$th cluster. Among each submap, a winner is computed (eq 2):

$$winner_i = max_{k \in n}(Y_i^k) \quad (2)$$

Like in SOM algorithm, the learning is made according to the topology of the submap (eq 3):

$$W_{kj} = W_{kj} + \epsilon \cdot Y_i^k \cdot \delta(d(winner_i, k), Pn, Nn) \cdot Z_i \quad (3)$$

The $d$ function computes a simple distance between the $k$th neuron and the winner neuron of the submap. The $\delta$ function computes the values of the lateral excitatory/inhibitory connections modulating the learning of the winner neuron's neighborhood. $\delta$ is a difference of gaussian (DOG) function whose shape is generated by the positive and negative neighborhood values ($Pn$ and $Nn$, see figure 7 for more details). $\epsilon$ is the learning rate.

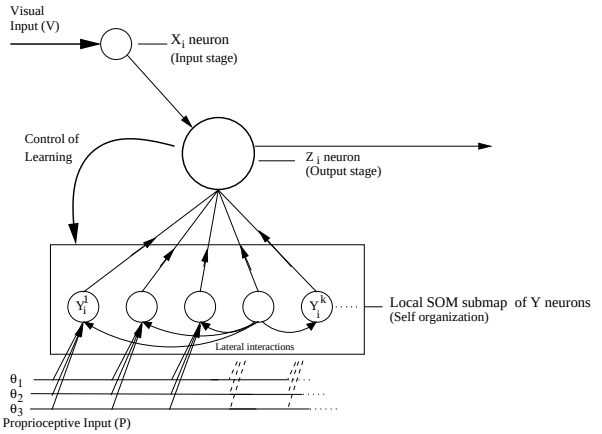


Figure 6: Representation of a cluster of neurons (here the $i$th cluster). Each cluster has one link with one neuron of the visual map. This link carries out visual inputs detected by $X_i$. Activation of $X_i$ triggers $Z_i$ and the learning of the associated $submap$. The learning consists in a self-organization of the $Y_i^k$ neurons associating proprioceptive signals($\theta_1$, $\theta_2$, $\theta_3$ of the arm) to one visual information.

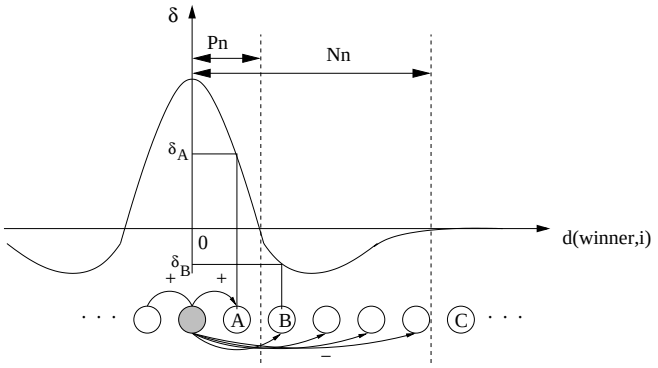


Figure 7: The $\delta(winner_i, k, Pn, Nn)$ function is a classical DOG. The winner neuron is in grey. $\delta$ is positive for neuron $A$, negative for neighbors $B$, null for the neurons $C$

The learning of a submap is dependant of the activation of the corresponding $X_i$ neuron, triggered by a visual input. (eq 4). Thus, each submap learns the different proprioceptive configurations independently.

$$X_i = \begin{cases} 1 & if\ V_j \cdot W_{ij} > \theta \\ 0 & otherwise \end{cases} \quad (4)$$

If no V is present, P is enough to trigger the response of the associated cluster in the map. The activity of the winner of each submap is then propagated to the potential of each associated $Z$ neuron (eq 5).

$$Z_i' = max(X_i, winner_i) \quad (5)$$

A simple competition process is then performed on the output $Z$ neurons of the map (eq 6).

$$Z_i = \begin{cases} 1 & if\ Z'_i = max_{j \in n}(Z'_j) \\ 0 & otherwise \end{cases} \quad (6)$$

The winner neuron $Z$ will represent the "visual' response associated to the proprioceptive input presented. Thus, many proprioceptive configurations are able to activate the same "visual feeling", while close visual responses can be induced by very different proprioceptions (thanks to the independence between each cluster).

### *3.3 On-line learning*

During the learning phase, the robot is put in a "quiet" environment, far from possible ambiguous distractors (learning with distractors would require much more presentations to detect the stable part of the sensory-motor associations). A crucial part of the learning process is linked to the choice of the learning parameters. Learning is uniquely controlled by the "shape" of the lateral inhibition between $Y$ neurons of a submap ($\epsilon$, $Nn$, $Pn$, involved in eq 3). This will influence the coarse to fine process which is mainly represented by two different stages: **A coarse stage:** at the beginning of the learning process, each cluster learns one visuo-motor association, self-supervised by vision (fig 8.b, 8.c). $\epsilon$ is maximal, $Pn$ high, there is no lateral inhibitions (the shape of the lateral interaction is a positive Gaussian). At this stage, the system does not cope with multiple arm positions for one visual position. The robot can perform either random or continuous arm movements, since the topology is forced by the presence of an V signal. The learning parameters are: $\epsilon = 0.9$, $Pn = 7$, $Nn = 7$. Figure 8.c) required 300 iterations of learning.

**A tuning stage:** The system will learn the equivalences between multiples different positions of the arm and a single visual position of the extremity (fig 8.d, 8.e, 8.f). It consists in a dissociation phase of the neurons of the submap, whose lateral inhibition shape and amplitude are diminishing thru time (progressive stabilization). The system progressively copes with multiples arm positions. The robots has to perform numerous random arm movements, to trigger each cluster with different arm positions. The learning parameters are: $\epsilon = 0.4$ down to 0.05, $Pn = 7$ down to 1, $Nn = 7$ down to 3. Figures 8.g) required 8000 iterations of learning.

Theoreticaly, the fine learning can work on its own, but it requires a very long time of convergence. "Fine" also refers to the learning constant which are in this case small, to allow a slow, progressive but accurate modification of the weights. Starting with the fine stage would only takes a long time before all the clusters would be separated.

Practically, and especialy in the case of learning robots, a coarse stage induce a large categorization of the space quickly: each cluster is easily separated rom others, and the robot is already able to perform coarses but consistent movements.

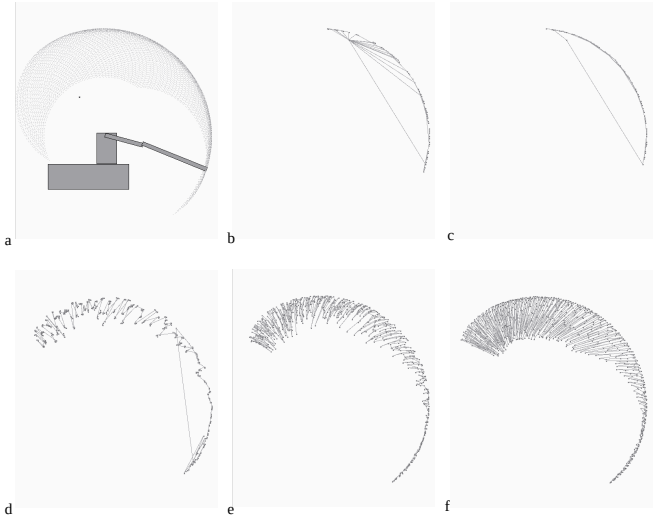


Figure 8: A sensory-motor vector of 146 clusters learning vertical movements (only the $\theta_2$ and $\theta_3$ joints of the arm where freed). Each cluster is composed of a self-organizing submap of 6 $Y$ neurons. a): Representation of the arm, and the theoretical working space. Each point is an accessible positions of the extremity of the arm, to be learned. On b)c)d)e)f) examples, each point represents a position learned by an $Y$ neuron. These points are plotted according to the values of the $Y$ neuron's weights learning the $\theta_2$ and $\theta_3$ values, using a simulation of the robotic arm. b) and c): the coarse stage. d), e), f): Progressive dissociation of $Y$ neurons during the tuning stage.

## 4. An unique behavioral dynamic

To reach a perceived target, the error between the desired position (the visual position of the target) and the current position of the device has to be minimized. This error has then to be converted in an appropriate movement vector to move each device's joint toward the target. This process is done by computing a dynamical attractor centered on the target stimuli. The attractor is computed on two 1-D maps of dynamical neurons preserving the visual map topology, using neural field equations (eq 7, (Amari, 1977)):

$$\tau \cdot \frac{f(x,t)}{dt} = \begin{array}{l} -f(x,t) + I(x,t) + h \\ + \int_{z \in V_x} w(z) \cdot g(f(x-z,t))\, dz \end{array} \quad (7)$$

Without input, the homogeneous pattern of the neural field, $f(x,t) = h$, is stable. The inputs of the system, $I(x,t)$, represent the stimuli which excite the different regions of the neural field and $\tau$ is the relaxation rate of the system. $w(z)$ is the interaction kernel in the neural field activation. These lateral interactions

("excitatory" and "inhibitory") are also modeled by a DOG function. $V_x$ is the lateral interaction interval. $g(f(x,t))$ is the activity of the neuron $x$ according to its potential $f(x,t)$. This robust and dynamic representation allows to get for free the following properties (see (Moga and Gaussier, 1999) for experimental results on the use of NF as control architecture for autonomous robot):

- The bifurcation properties of the equations allow a reliable decision making if multiples stimuli are presented.
- The time constant induces a remanent activity of the neural field, proportional to the intensity and the exposure time to the stimulus. This memory property is a robust filter of non stable or noisy perceptive stimuli.

Moreover, the spatial derivate of the NF activities is interpreted as the two desired speed vector (horizontal and vertical), to reach the attractor target (*read-out* mechanism (Schöner et al., 1995)). According to their proprioceptive position, each joint will move at the value read on the derivation of the NF activities. Each joint will then contribute to the global move of the arm toward the visual objective, the shape of the NF activity on the target ensuring convergent moves. This coding of movements induces the following properties:

- The simultaneous contribution of each joint to the movement, is an emergent property of the architecture and the dynamic representation of the target.
- A smoothed speed profile with acceleration and deceleration phases of the joints at the beginning and end of the movement.
- The stabilization of the motors on the target ($d\phi = 0$).

The computation of NF equation on the perceptive activity allows a generic coding of the internal dynamic, independent of the motor devices, without taking into account the number of DOF of the device, the possible redundancies (this part being managed by the visuomotor learning), and therefore the possible mechanical changes that can be made during a robot's "life". Hence, the association of an adaptive sensory-motor map with two 1-D neural fields can be seen as a simple and global dynamical representation of the working space controlling efficiently an arbitrary number of degrees of freedom according to 2-D information coming from the visual system.

## 5. Results

After the learning phase, the architecture was tested in pointing and low-level imitation tasks. The pointing task is aimed to test how the internal dynamic (the *read-out* mechanism), and the coherency of the learned associations successfully drive the extremity of the arm to a desired visual area in the working space. The imitative task is aimed to test the robustness of the dynamical equations and the real time capabilities of the overall architecture. The imitative experiment also validate theoretical works assuming that a generic controller is able to perform imitation of human gestures (Gaussier et al., 1998, Andry et al., 2001).

### *5.1 Pointing experiment*

Figures 9 and 10 show the results of one pointing test, performed on one of our robots. To simplify the plotting of the results, the pointing was made using two DOF of the arm ($\theta_2$, and $\theta_3$). Nevertheless, the pointing test preserves the complexity of the issue since these two DOF are redundant in their contribution to the vertical movements of the end of the arm. The pointing test was

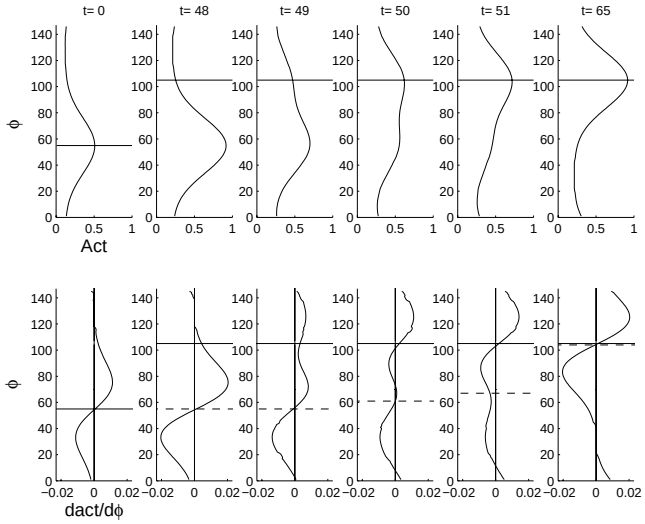


Figure 9: Internal activities of the neural field map and read-out mechanism during a pointing task. From $t = 0$ to $t = 47$ the neuron 55 is stimulated. From $t = 48$ to $t = 65$, the neuron 105 is stimulated. Up: snapshots of the neural field's activity. A single stable attractor can be seen at $t = 47$ and $t = 65$. Between $t = 47$ and $t = 65$ the attractor travels from the first stimulated point (neuron 55) to the second one (neuron 105). Bottom: corresponding activity of the NF spatial derivative. From $t = 47$ to $t = 65$, the resulting read-out mechanism moves the arm toward the new attractor (dashed lines represent the visual position of the extremity of arm, and $dact/d\phi = 0$ represent a null speed of the associated joint).

performed as follow:
Vision from the CCD camera is disabled (no visual information), and the activity of the neural field is directly controlled by the experimenter, simulating a perfect perception. At the beginning of the experiment (from $t = 0$

to $t = 47$), there is no error between the activated area (centered on neuron 55) of the neural field and the position of the arm, inducing no movements. At $t = 48$, the experimenter stops the activation of the neuron 55, and starts to stimulate the neuron 105. The resulting traveling wave (from $t = 48$ ) on the NF's activity (fig 9,up) induces modifications of the shape of the associated spacial derivative (fig 9, bottom), and starts to move (from $t = 49$ ) the arm's joints toward the new equilibrium, progressively centered on the neuron 105. Figure 10 shows a plot of the recorded positions of the arm corresponding to the modifications of the NF's activity between $t = 48$ and $t = 65$ according to two modalities. In the first modality, the $\theta_2$ joint was blocked, and only the $\theta_3$ joint was able to point in the direction of the target (fig. 10.1). In the second modality both joints $\theta_2$ and $\theta_3$ where freed and able to reach the target (fig. 10.2). These records show that in both modalities, successful pointing is achieved by the system. We can notice that: first, the pointing is achieved with minimal precision error inferior to 3 degrees, (with a resolution of 1.5 degrees of the neural network coding), second the sensory-motor map has learned a reliable visuo-motor space (no incoherent or discontinuous movements), and third both commands correspond to a single internal dynamic. These examples of pointing show that the architecture exploits efficiently 2-D visual informations for positioning the robotic arm in the 3-D surrounding space. The 2-D to 3-D position equivalence of the extremity of the arm is performed thanks to the previous learning of the sensory-motor map, which topology favor minimal cost movements by minimizing the distance of each joint between the current proprioceptive configuration and the desired visual target position of the extremity. To sum-up, the visually forced topology of the sensory motor map ensures the shortest path of the extremity of the arm to reach the target, while dynamical neural fields ensure the proper speed profile of the extremity of the arm, and the contribution of all the aviable joints to the whole movement whatever the number of joints used is.

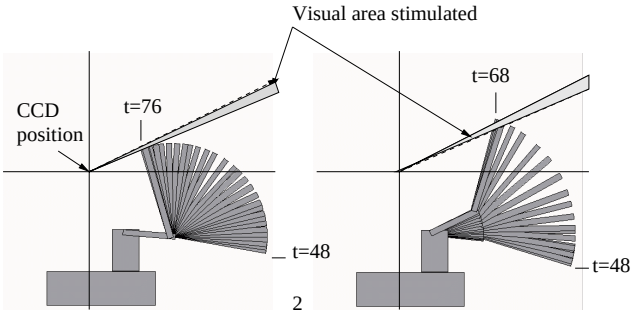


Figure 10: Pointing results, resulting of the NF activity plotted in figure 9. Left: one DOF modality. Right two DOF modality. In both examples, the same internal dynamic succeeds to drive the arm to the area of the visual space stimulated.

## *5.2 A Contribution to imitation behaviors*

In (Gaussier et al., 1998) we showed that a mobile robot is able to imitate successions of displacements performed by a human demonstrator with a control architecture based on the two following principles:

1. The perception is fundamentally ambiguous
2. The robot is nothing more than an *homeostat*. The robot tends to reduce the error between its visual perception and its proprioception (it can be seen as visuo-motor reflex allowing to follow the teacher direction).

From these generic principles, the imitative behavior can be seen as a side effect due to the perception limitation of the system: walking in front of the robot induces changes in the perceptions that the robot interprets as an unforeseen self movement. It tries to correct by opposite movements, inducing the following of the demonstrator. In the present system, the generic principles of the architecture are the same:

- same perception system
- the homeostatic behavior is the core of the control of arm movements (a difference between visual stimulation and motor positions induces the movement of the arm toward the visual stimulation, as demonstrated by the pointing task).

Moreover the learning process constitutes a primary behavioral repertory. Thus, an elementary imitative behavior can be triggered by exploiting the *ambiguity* of the perception. Using only movement detection, the system can't differentiate it's extremity from another moving target, such as a moving hand. By shifting the head horizontal motor with the robot's body proprioception, we ensure that the robot's arm can't be in the field of view of the camera. Thus, a perceived moving hand will be associated to the robot own arm (perception ambiguity). The generated error will induce movements of the robotic arm reproducing the moving path of the human hand: an imitative behavior emerges. Using this setup, we show that our robot can imitate several kind of movements (square or circle trajectories, up and down movements, see fig 11). During the experiment, the 3 main DOF of the arm where freed, allowing movements in all the working space. The experimenter was naturally moving its arm in front of the robot's camera making simple vertical or horizontal movements, squares, or circles. The camera rapidly tracked the hand (the most moving part of the scene) and the arm, reproduced in real time

the hand's perceived trajectory. The use of neural fields ensures a reliable filtering of movements and a stable, continuous tracking of the target by the head and the arm of the robot.

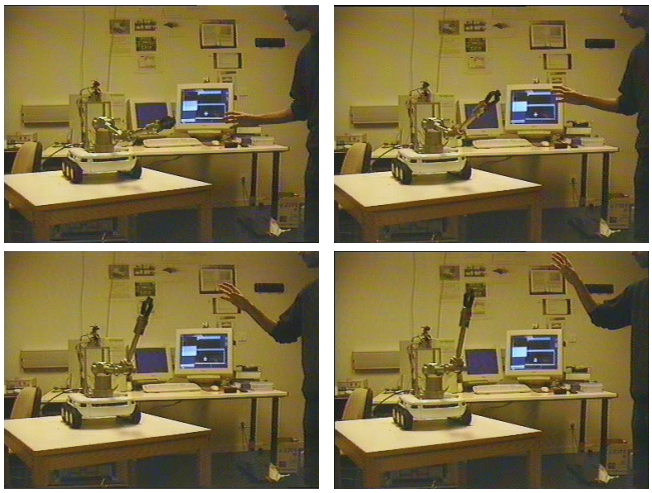

Figure 11: Real time imitation of simple vertical gesture. To obtain a low-level imitative behavior, we simply shift the head's position with the body and the arm ($shift = 90$ degrees). Thus, a perceived movement is interpreted as an error, inducing corrective movements of the arm: an imitating behavior emerge

## 6. Conclusion

Far from building a control architecture dedicated to imitation tasks, we showed how a generic system with learning capabilities and dynamical properties, can easily exhibit low-level imitations. These imitations of arm movements are performed without any internal model of the human arm, and can easily be transposed to imitation of robotic arm movement, whatever the morphology of the arm is. We are here close of an *effect level imitation* (Nehaniv and Dautenhahn, 1998), where real-time executions, and dynamics of the movement are suffisant to provide low-level but efficient imitation (efficient for interaction, because imitation of movements can be recognized, and efficient for learning, because information about the most useful and informative part of the movement, the end point, is used).

We have also shown how a very simple vision system, only built on movement recognition, is sufficient to obtain a robust end point tracking behavior. We have also shown that the on-line learning of the coordination of complex robotic devices can be solved by an appropriate design of the starting architecture (The choice of a same topological coding for vision and motor control). Indeed, the low-level imitative behavior constitutes the trigger of the next developmental stage. It allows gestural interaction with humans, objects, and the "outside world". These interactions constitute a new way for the system to learn more complex sensory-motor associations, about the physical, dynamical and also social properties of the environment. For example, we plan to make our robot learn sequences of arm movements (already tacked in (Gaussier et al., 1998) with a mobile robot) via interaction with humans or robots. We are investigating learning issues allowing our architecture to learn, store and reproduce sequences of gestures, but also to detect implicit but important information during a gestural interaction. From a developmental point of view, we assume that higher capacity of interactions will be built up from experiencing more and more complex situations, carried out by the acquisition of new sensory-motor capabilities.

## References


Amari, S. (1977). Dynamic of pattern formation in lateral-inhibition type by neural fields. *Biological Cybernetics*, 27:77–87.

Andry, P., Gaussier, P., Moga, S., Banquet, J., and Nadel, J. (2001). Learning and communication in imitation: An autonomous robot perspective. *IEEE transactions on Systems, Man and Cybernetics, Part A*, 31(5):431–444.

Billard, A. (2001). Learning motor skills by imitation: A biologically inspired robotic approach. *Cybernetics and Systems: An International Journal*, (32):155–193.

Bullock, D., Grossberg, S., and Guenther, F. (1993). A self-organizing neural network model of motor equivalent reaching and tool use by a multijoint arm. In *Journal of Cognitive Neuroscience*, volume 5, pages 408–435.

Cheng, G. and Kuniyoshi, Y. (2000). Real time mimicking of human body motion by a humanoid robot. In *Proceedings of The 6th International Conference on Intelligent Autonomous Systems (IAS2000)*, pages 273–280.

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.

Hainline, L. (1998). The development of basic visual abilities. In Slater, A., (Ed.), *Perceptual Development*, pages 5–50. Hove: Psychology Press.

Kohonen, T. (1976). *Associative memory, a system-Theoretical approach*. Springer-Verlag.

Marjanović, M., Scassellati, B., and Williamson, M. (1996). Self-taught visually-guided pointing for a humanoid robot. *Proceeding of the Fourth International Conference on Simulation of Adaptive Behavior*.

Moga, S. and Gaussier, P. (1999). A neuronal structure for learning by imitation. In Floreano, D., Nicoud, J.-D., and Mondada, F., (Eds.), *Lecture Notes in Artificial Intelligence - European Conference on Artificial Life ECAL99*, pages 314–318, Lausanne.

Nadel, J. and Butterworth, G., (Eds.) (1999). *Imitation in Infancy*. Cambridge: Cambridge University Press.

Nehaniv, C. L. and Dautenhahn, K. (1998). Mapping between dissimilar bodies: Affordances and the algebraic foundations of imitation. In Demiris, J. and Birk, A., (Eds.), *Proceedings of Seventh European Workshop on Learning Robots 1998 EWLR98*, pages 64–72. World Scientific Press.

Niemeyer, G. and Slotine, J. (1988). Performance in adaptive manipulator control. In *International Journal of Robotics Research*, volume 10.

Schaal, S., Atkenson, C., and Vijayakumar, S. (2000). Real time robot learning with locally weighted statistical learning. In *International Conference on Robotics and Automation*.

Schöner, G., Dose, M., and Engels, C. (1995). Dynamics of behavior: theory and applications for autonomous robot architectures. *Robotic and Autonomous System*, 16(2-4):213–245.

Vinter, A. (1985). *L'imitation chez le nouveau né*. Neuchatel - Paris, delachaux and niestlé edition.