Human-Robot Interactions as a Cognitive Catalyst for the Learning of Behavioral Attractors.

C. Giovannangeli and P. Gaussier
ETIS Laboratory - CNRS UMR 8051
Neurocybernetic Team
Cergy-Pontoise University
e-mail: giovannangeli@ensea.fr

***Abstract*— We address in this paper the problem of the autonomous online learning of a sensory-motor task, demonstrated by an operator guiding the robot. For the last decade, we have developed a vision-based architecture for mobile robot navigation. Our bio-inspired model of the navigation has already proved to achieve sensory-motor tasks in real time both in unknown indoor and outdoor environments. We propose to bootstrap the underlying PerAc architecture in order to control the sensori-motor learning. The interaction leads the robot to autonomously build a precise sensory-motor dynamic approximating the behavior of the teacher. A real dialog based on actions imposed by the teacher and those proposed by the robot emerges, which catalyzes the learning of the robot. The architecture is finally tested in real indoor and outdoor environments.**

I. Introduction

Task specification in autonomous robotics has aroused an increasing interest for about ten years. It is today admitted that autonomous mobile robots should be designed without prior knowledge of the tasks to perform. On the contrary, they should be designed to constitute their skills via interactions with their physical and social environment where they accumulate experience of the sensory-motor world [1], leading their own cognition to enact [2]. In this context, Human-Robot Interactions (HRIs) are thought to be a very efficient means to bootstrap the learning of a robot or to specify some various tasks to a robot [3]. HRIs also appear as a key point for designing social and adaptive robots [4]. More generally, autonomous navigation implies control constraints, machine learning constraints and more recently HRI constraints for the applications involving humans.

This paper aims at illustrating how a trainer can intuitively condition an autonomous robot to follow a route or to perform a homing behavior. We propose a visual navigation architecture bootstrapped for task specification by HRIs, which provides an efficient autonomous control system for mobile robots engaged in patrol or exploration missions. The guidance of the robot by a joystick will be used as a simplification of a process of training: the robot can be seen as a dog we want to train by pulling or releasing its leash. In [5], the problem of the task specification is approached as the estimation of a sequence of concurrent behaviors already mastered by the robot (the behaviors have been previously learned). The authors also insist on the fact that *acting* can provide a basis for a non-verbal human-robot communication and appears as a simple means for the robot to exhibit that it requires some help from the trainer. The idea that the robot could ask questions to its trainer has already been evaluated, for example in the collaborative control of [6]. "The robot ask questions to the human..." which are translated into a comprehensive human language "... in order to obtain assistance with cognition and perception". The answers are translated in the symbolic language of the robot. As a general rule, task specification and communication are performed at a very high symbolic level under the dictatorship of the trainer. A less supervised process for task specification could emerge from a non-verbal interaction. However, most of the robotic architectures using for instance imitation need to separate the learning phases and the performance phases. It is contradictory with lifelong learning constraints which imply that the robot must be able to learn while freely moving in the world. In the context of imitation, learning and demonstration phases should be gathered to provide a rich and natural communication improving the development of the robot skills: by imitating a professor, the robot experiments the behavior that has to be learned. By *acting* and *reacting* to the trainer orders, the robot freely exhibits its mastery of the task while in parallel learning the task. Although it is a non verbal, non symbolic communication, it is nevertheless a rich communication, which will be shown to catalyze the learning of the robot.

This paper first presents the visual system of the robot enabling to create a continuous spatial state space. We propose an architecture that enables the semi-supervised learning of a sensory-motor behavior (a route, a homing behavior). The robot refines a sensory-motor dynamic that approximates the desired behavior. The system does not require the separation of learning and performance phases, which are scattered in the time according to the rhythm of the interaction and to the sensory-motor error. Using simulations, we show how the human being enriches the interaction, as compared to a simulated teacher or an ad-hoc process. HRIs enable a more accurate approximation of the desired behavior. Finally, visual navigation experiments in real indoor and large outdoor environments are presented. The real time features of our algorithms provide an attractive and intuitive HRI which catalyzes the learning of the robot, and reveal its efficiency for the interactive learning of complex navigation tasks.

II. PERAC ARCHITECTURE FOR VISUAL NAVIGATION

Among the various methods to create spatial behaviors, the PerAc architecture [7] has demonstrated to be particularly adapted for sensory-motor learning. A PerAc architecture can underly many tasks in mobile robotics: guidance, local navigation in indoor [8] and outdoor environments [9], reproduction of a temporal sequence of actions, as well as in the control of multiple freedom degrees actuators: arm robot control [10], gaze direction control. This architecture is able to learn sensory-motor associations which occur during the sensory-motor life of the robot. This paper focuses on the PerAc architecture for local navigation. A model of visual place-cells is used to provide a robust localization level of the robot in indoor as well as in outdoor environments [11]. Here, we address a more general class of algorithms based on state recognition, which can lead to an adaptive state space partitioning (environmental paving). For example, systems based on GPS measurements, triangulation systems via external references, classical SLAM, vision-based SLAM or topological approaches of SLAM could produce the primary data for the algorithms we present.

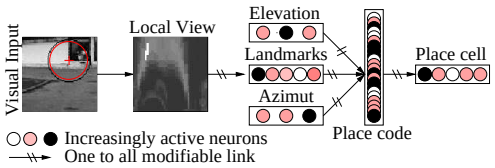


Fig. 1. Block diagram of the place recognition architecture: it is composed of a visual system that focuses on points of interest and extracts small images in log-polar coordinates (called local views), a merging layer that compresses *what* and *where* information, and a place recognition layer.

Fig. 1 summarizes the processing chain used on our robots for place recognition. A place is defined as a spatial constellation of on-line learned visual features (here a set of triplets *landmark-azimut-elevation*) compressed into a place code. The constellation results from the merging of a *what* information and a *where* information provided by the visual system that extracts local-view[1] centered on points of interest. The *what* and *where* merging implies that the shape of the place field is homothetic with the shape of the environment [12], [11] (*i.e* the place fields extend with the distance to the landmarks). Moreover, neither Cartesian nor topological map building is required for the localization. On the contrary, the world acts as an outside memory [13] of the static invariants of the attentional vision system [1]. Inasmuch as the learned invariants of a location persist in its neighborhood, the localization is possible without map building.

A simple associative learning between places and actions enables to generate a sensory-motor attraction basin for homing or path following behaviors. The problem of the self-constitution an efficient policy of action has often been stressed in the literature of reinforcement learning [14] but we claim that the PerAc architecture is extremely

[1]A log-polar mapping is used to transform these local view, providing some robustness to scale and rotation variation.

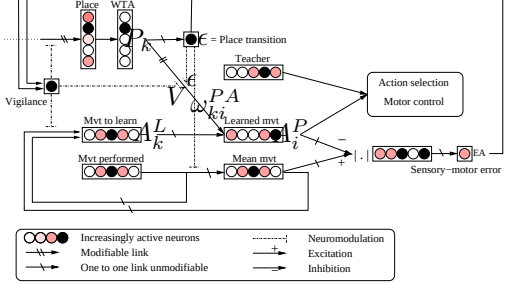


Fig. 2. Bootstrapped PerAc Architecture: A one shot learning enables the one-shot learning of places and place-action associations and the refinement of the sensory-motor dynamic. Computation of a signed angular error between the movement and the predicted movement in a given place enable to adapt movement associated to this place. The one-shot learning of landmarks, constellations, places and place-action associations is triggered by a vigilance signal, whereas the adaptation is performed continuously.

efficient for spatial behavior learning since the problem of the state space partitioning as well as the problem of policies learning are embedded together. The next section presents a generic bootstrap for the PerAc architecture enabling the autonomous online building of an attraction basin around a behavioral attractor during an interactive learning.

III. REFINEMENT OF A BEHAVIORAL DYNAMIC

We propose here a control architecture enabling an autonomous mobile robot to learn and to reproduce a sensory-motor tasks while interacting with a human (autonomous exploration and reinforcement learning could also be classically used, but the time of learning is prohibitive for the kind of applications we interest in). Being guided by the human, the robot learns places and is able to merge the action associated to the current state (here places) with the action imposed by the trainer. We use a joystick to guide the robot but a visual tracking of the trainer could also have been possible[2]. The bootstrap we propose allows the PerAc architecture to learn in one shot novel place-action associations and to adapt the associated movements according to the sensory-motor error generated during the crossing of this place. We specify the sensory-motor neural network (see fig. 2) and the equations underlying the control architecture.

A. Refinement of a sensory-motor dynamic

The neural group performing the sensory-motor learning is inspired from the Least Mean Square neural network (LMS)[15]. The neural architecture is given in fig 2. The algorithm for updating the activity of the neurons performing the sensory-motor learning is:

$$s_k(t) = \sum_{i=1}^{N_P} \omega_{ik}^{PA}(t) P_i(t) \quad (1)$$

$$A_k^P(t) = V(t).S_k^d(t) + (1 - V(t)) \cdot \left(\frac{s_k(t)}{s_{max}(t)} \right) \quad (2)$$

[2]This kind of interaction is an unidirectional *imitation process* so we can not talk about pure imitation. Indeed imitation implies that each agent can imitate the other (role switching), which does not occur in the proposed interaction.

In this equation, $S_k^e(t) = A_k^e(t)$ is the activity of the kth input neuron, which provides either the previous performed movement when the vigilance spikes (enabling the one-shot learning) or the mean movement since the last place transition (enabling a delayed adaptation). The mean movement is reseted by the $\epsilon(t)$ signal as shown in fig. 2, each time a place transition occurs. $s_k(t)$ is the predicted activity of the kth neuron of the group. $P_i(t)$ is the normalized activity of the most activated place cell i: $P_i(t) = 1$ if the current place is the place i and $P_i(t) = 0$ otherwise. ω_{ik}^{PA} is the weight of the connection between the ith place cell and the kth action. Finally, $s_{max} = \max_{k=1..N_{Ac}}(s_k)$ is used for the output normalization (N_{Ac} being the number of neuron coding an action). More precisely, $S_k^d(t)$ is the desired output (the future action to predict explicitly given by the input group called *Mvt to learn* in fig 2.), the equation 1 computes the predicted output and the equation 2 provides the output of the group computed either as the normalized prediction or as the desired output during a one-shot learning cycle (no prediction being available before the one-shot learning).

We consider two signals for the bootstrap of the sensory-motor learning. The first signal is a vigilance signal $V(t)$ which triggers the waves of one-shot learning (a one-shot learning cycle during which a new place-action couple is learned). The second signal $\epsilon(t)$ corresponds to a learning rate. It is used as a modulation for both the one-shot learning and the long term adaptation. The main difference with a LMS is that the weight modifications are composed of two terms. A term performing a one shot learning computed as the classical gradient of a Widrow-Hoff (WH) algorithm and a term computed according to the previous gradient computation, corresponding to a delayed learning. In our architecture, $\epsilon(t)$ spikes each time a place transition occurs (hence also each time each time the vigilance signal spikes). The update of the synaptic weights is performed after the update of the neurons activity according to the following equations:

$$G_{ik}^i(t) = (S_k^d(t) - s_k(t)).P_i(t).V(t) \quad (3)$$

$$G_{ik}^d(t) = (S_k^d(t) - s_k(t)).P_i(t).(1 - H_0(\epsilon(t))) \quad (4)$$

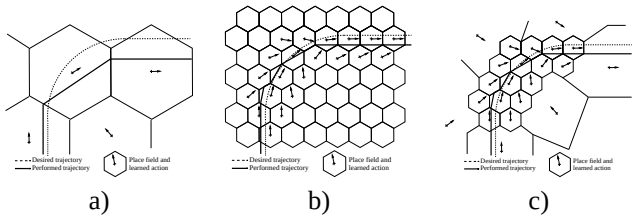
$$\frac{d\omega_{ik}^{PA}}{dt} = (G_{ik}^i(t) + G_{ik}^d(t - dt)).\epsilon(t) \quad (5)$$

In this equation, two gradient terms are computed: G_{ik}^i (instantaneous gradient) which is the classical WH gradient with a term of vigilance that modulates the learning and G_{ik}^d (delayed gradient) which computes a gradient if no learning or adaptation occurs (H_x being the Heaviside function: $H_x(y) = 1$ if $y > x$ and 0 otherwise). During a one-shot learning cycle (when the $V(t)$ and $\epsilon(t)$ spike), the new place is associated with the current action (classical WH learning) by means of the not null terms $G_{ik}^i(t)$ in the equation 5. In the general case, a delayed adaptation is performed each time $\epsilon(t)$ spikes by means of the term $G_{ik}^d(t - dt)$ (the previous gradient). In our case, $\epsilon(t)$ corresponds to a place transition (it can be expressed as $\epsilon(t) = [P_i(t) - P_i(t - dt)]^+$, with with $[x]^+ = x$ if $x > 0$). Hence, the adaptation of the movement in a place is performed only once the robot has left the place and will be available only the next time the robot will re-enter the place. As a general rule, the adaptation of a sensory-motor association requires a kind of learning evaluation and can only be performed after the sensory-motor association has occurred. In the context of the sensory-motor learning, this delayed adaptation seems to be crucial to control the instants of learning.

This section proposed a bootstrapped version of the PerAc architecture enabling a robot to refine a sensory-motor dynamic at the motor level. The remaining question concerns the control of the vigilance signal: Which signals are important for the autonomous partitioning of the state space corresponding to a refinement at the sensory level?

B. Autonomous and adaptive state space partitioning

In the context of the reproduction of a trajectory, the important criterion is the precision of the reproduced trajectory which is directly linked to the spatial discretization of the behavioral dynamic. Paving the environment according to a single recognition threshold on the place cell activity (as in our former model) implies that the size of the place field and so the precision of the spatial encoding are fixed (see fig. 3 a) et b)). Since the regularity of the paving imposed the precision of the behavioral dynamic, the paving of the environment should not be regular but adapted to the desired precision and the complexity of the trajectory (see fig. 3 c)). For instance, more place-action associations should be encoded during a sharp bend than during a straight line. The system has to use the discrimination capability of the place recognition in the complex parts of the trajectory and its generalization capability in the easier. In a more general context than the navigation, the assumption that a sensory-motor dynamic $D : S \rightarrow M$ is better approximated if the discretization factor of the sensory space S evolves as the variation $\frac{dD}{dS}$ remains valid (the compression factor is adapted to the variations of information).



a) b) c)

Fig. 3. Comparison of a regular partitioning (left and center) versus an adaptive partitioning (right) of the state space. The precision of the reproduced trajectory depends on the precision of the state space partitioning (the size of the cells). The adaptive partitioning try to find a trade-off between the precision and the number of encoded place-cell, whereas a higher precision implies more encoded place-cell with a regular partitioning.

The sensory-motor error E_a in each place can be defined as the difference between the predicted and the performed action. It stands for the parameter $\frac{dD}{dS}$: indeed, the sensory-motor error is higher in complex parts of the trajectory than in easier parts, because more changes of direction occur. Hence, the sensory motor error appears as a pertinent signal to control the learning of the places. An angular threshold $t_{E_a}^+$ on the sensory-motor error is responsible for the accuracy of the behavior during a bend. For example, if

$t^+_{E_a}$ is 30°, then a 90° bend should be encoded by at least three place-cells. Thus, this measure can be used to control the vigilance signal in order to create a state space adapted to the complexity of the desired behavior. In order to avoid the over-coding of the environment, a safety threshold t^+_P over which a place is considered as recognized can be fixed. If the maximal place-cell activity P_M is higher than t^+_P, the coding of new place-cells is inhibited. This threshold can be as high as the discrimination capability of the place recognition. We finally use a low threshold t^-_P to trigger the learning of a new place when all the other encoded places are not enough recognized. t^-_P must be correlated with the generalization capability. Thus, the vigilance signal is defined as:

$$V = H_0 \left(\ (t^+_P - P_M) \ \times \ ([E_a - t^+_{E_a}]^+ \ + \ [t^-_P]^+ - P_M) \ \right)$$

The use of the sensory-motor error E_a to control the learning of a new location allows to adapt the precision of the spatial partitioning to the curvature radius of the trajectory (equivalent to the difficulty). Moreover, precise thresholds do not have to be estimated, but confidence thresholds for the recognition and the non-recognition. The following sections propose experiments to illustrate the pertinence of this architecture, especially in the context of task specification by HRIs.

IV. HRI as a Cognitive Catalyst

This section evaluates in a simulated environment the role of the HRIs in the learning of a sensory-motor dynamic, consisting here in reproducing an arbitrary trajectory. We propose two measures in order to compare the optimal trajectory $\{x_i(p)/p \in \{1..P\}\}$ with the reproduced trajectory $\{x_r(t)/t \in [t_i..t_f]\}$:

$$e_t = \int_{t=t_i}^{t=t_f} \min_{p=1}^{P} \|x_r(t) - x_i(p)\|.dt \tag{6}$$

$$e_p = \sum_{p=1}^{P} \frac{\min_{t=t_i}^{tf} \|x_r(t) - x_i(p)\|}{P} \tag{7}$$

A combined measure may also be used, such as $(e_t + e_p)$.

Two strategies of teaching can be simulated: A prescriptive teaching and a proscriptive teaching. On one hand, the prescriptive teaching consists in a perfect guiding of the robot without observing its behavior, as shown in the first experiment of the fig. 4 (this kind of teacher is closer to a demonstration than a guidance process). Since no interaction occurred, and no error had been committed, the algorithm is not able to efficiently generalize. Indeed, the created dynamic is not stable: the dynamic can either lead to a behavioral attractor (but the generated trajectory is quite unsatisfying), or lead to a second parasitic attractor (in the center of the environment), or lead the trajectory to diverge. On the other hand, the proscriptive strategy consists in only correcting the robot when it is too far from the center of the trajectory, according to a given threshold (second experiment of the fig. 4). This strategy has the advantage for the trainer to evaluate directly the precision of the learning by observing the errors of the robot. Moreover the locations of the place-action associations surround the precise trajectory, allowing the convergence of the different trajectories through an attractive trajectory, defined by the points which are at the same distance from the two closer sensory-motor couples. Moreover, fig. 4 shows that the size of the attraction basin is largely wider than the experimented area. Yet, the robot difficultly stabilizes on the precise trajectory but oscillates around it.

In the last experiment of fig. 4, a human trainer presses a button when he wants to correct the behavior of the robot. The human and the robot really interact by means of a non-verbal, non-symbolic language based on actions (imposed by the trainer and reproduced by the robot). The natural course of the HRI, during which the teacher oscillates between a precise demonstration of the trajectory and observing-correcting the robot trajectory will produce both proscriptive and prescriptive teaching phases. As a result, the generated trajectories are more precise. The two strategies have actually complementary properties and occur successively during the real interaction. The proscriptive learning enables to create the border of the attraction basin, quite broadly but guaranteeing a convergence to the center of the trajectory, whereas the prescriptive learning enables to precisely model the behavioral attractor in the center of the trajectory. The evaluation of the generated trajectory shows that a human is more efficient than the two other ad-hoc teaching process (purely proscriptive or purely prescriptive strategies). The HRI has enabled to catalyze the learning of the sensory-motor dynamic.

The next section proposes experiments in real environments to highlight the usability and the accuracy of our *interactive* control architecture.

V. Experiments with Real Robots

The experiments proposed here show the accuracy of our approach in real environments. The indoor experiment (see fig. 5) demonstrates in favorable conditions that it is quite easy to train a robot to perform a sensory-motor task. Three laps were sufficient for the robot to adopt a convergent behavior. The precision could have been further enhanced by longer guiding the robot even when the spatial learning had converged. In the experimented environment ($\simeq 10 \times 7$m), the distance between the robot and the desired trajectory is never greater than 1.5 robot-widths of the robot (Koala KTeam robots are 30 cm wide). Outdoor experiments are far more difficult due to the constraints of the natural environments. To bound the effect of the non planar terrain, we developed a stabilized plate-form which maintains the camera in the horizontal plan. An automatic gain and exposition time adapter was necessary to reduce the effects of light changing (between a wall reflecting the sun and a wall in the shadow for example). However, we manage to teach the robot to reproduce a very precise looped trajectory, relatively to the expected theoretical precision, with only two laps of proscriptive learning (see fig. 5). Only 14 places were learned which is extremely cheap. The mean speed of the robot is about $0.4 \ m.s^{-1}$: the 200 meters trajectory is reproduced in 9 minutes.

A crucial problem of the interactions between humans and robots is that the human can never know if the task

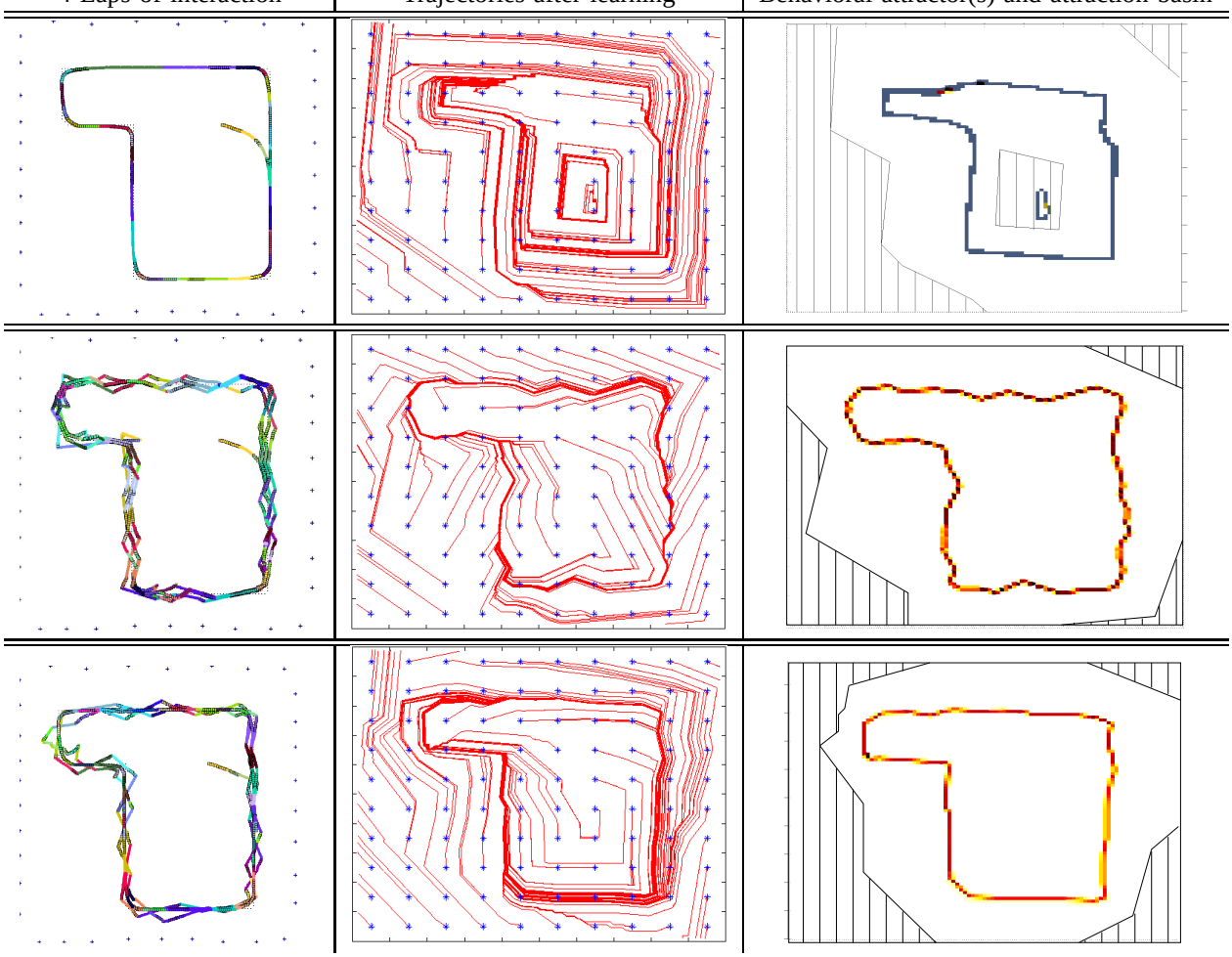


Fig. 4. Left figures show the trajectories during the four laps of training. The figures in the center show some generated trajectories. The behavioral attractors and their attraction basins are displayed on the right figures (the attractors correspond to the mean position of the robot after a long time for different starting points and the attraction basin is deduced from the fig. of the trajectories). In the first experiment, the prescriptive teaching is simulated. The trajectories either diverge or converge on a bad attractor. For this attractor, $e_t = 45$ and $e_p = 40$. A second parasitic attractor has also been created (bifurcation). In the second experiment, the proscriptive learning is simulated. The trainer never shows the precise trajectory to the robot. He only corrects the robot when it escapes too far from the desired trajectory according to a given threshold (here 20 pixels). As a result, the attraction basin is far wider. The robot oscillates around the desired trajectory but difficulty stabilizes on it. Only one attractor has been created. For the generated attractor, $e_t = 10.7$ and $e_p = 12.4$. The last experiment evaluates the human teaching. The human chooses when he wants to correct or guide the robot by simply pressing a button. The robot trajectories no longer bifurcate and the robot is able to precisely follow the desired trajectory. For the generated attractor: $e_t = 6.5$ and $e_p = 6.9$, which is the best score among the three experiments.

is really learned by the robot, and the robot never knows if the teacher is satisfied with its behavior. As none of them can evaluate the other, how is it possible for the robot or for the human to know that the task is learned? When the professor guides the robot, he is deprived from knowing what would have been the autonomous behavior of the robot. Hence, a single prescriptive learning is insufficient to produce a constructive interaction as previously illustrated. The professor also has to evaluate the robot behavior by proscriptively teaching it. The more the robot make errors, the more it learns. The actions and reactions of both the teacher and the robot leads a real dialog based on actions to emerge, which speed up the learning of the robot as compared to an ad-hoc teaching or a classical reinforcement learning algorithm: HRI appears as a cognitive catalyst.

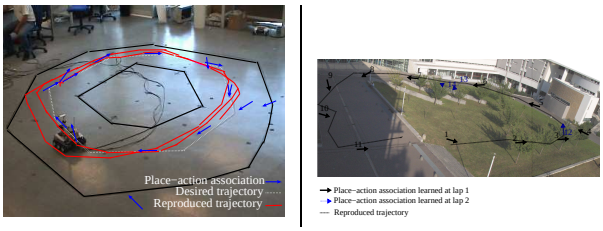


Fig. 5. Indoor experiment: the robot is guided by a human operator. Three laps are sufficient to train the robot to perform the task within the road defined by the black border. Outdoor experiment of the interactive learning of a visual path. After two laps of proscriptive learning (14 place-action associations learned), the 200 m are covered in 9 mn. Despite the architecture is split on four processors, the speed is limited ($0.4\ m.s^{-1}$) by the low level drivers of the robot. Otherwise the robot could operate faster.

VI. Conclusion and Perspectives

The goal of this paper was to present our results on human-robot interactions for navigation task specification.

We wondered how a robot can acquire some knowledges through a non symbolic and non verbal interaction with a teacher that specifies a dynamical task to the robot. This issues was addressed in the context of the autonomous navigation. A wired bootstrap enables the PerAc architecture for local navigation to adapt the partitioning of the state space to the complexity of the desired trajectory. A modified Widrow-Hoff learning enables the refinement of the sensory-motor dynamic by means of a delayed adaptation. The use of a joystick to teach the robot, in spite of its simplicity, creates a real interaction and leads to the emergence of a real *dialog* based on a vocabulary of actions (action imposed by the teacher and performed by the robot). The interaction is close to the training of a pet by means of a leash which is sometimes taut and sometimes not. Hence, the system does not separate the learning and performance phases, which unsupervisely alternate according to the rythm of the interaction. We pointed out the fact that human-robot interactions structure the learning and speed up its convergence. Finally, the experiments with real robots in indoor as well as in outdoor environments illustrated the efficiency of our solutions. Our works highlight it is really much more constructive to use proscriptive strategies than a prescriptive strategy to obtain good generalisation capabilities and stable dynamical behaviors.

Yet, controlling the end of the interaction remains a central problem. If the trainer is not satisfied with the robot behavior whereas the learning has already converged, this corresponds to the fact that the desired precision is not available for the robot which already performs the task as well as it can. On the contrary, even if the trainer is satisfied with the robot behavior, the robot may however know some states in which it can progress. It can communicate its desire to progress, or even guide the trainer in these states. The interaction could also be more constructive if the robot has the possibility to disobey the teacher in known states or expresses its need of help in non-mastered states. Such variations of the behavior would constitute another interesting feed-back for the trainer on the mastery of the task by the robot. The problem of the self-evaluation arises. *The robot has to know what it knows*: it should be able to know if its learning enables him to progress, or if its predictions are normal according to the current situation. Our future work will study a progress-based approach derived from [16], [17], for the self-evaluation and the metacontrol of the learning.

ACKNOWLEDGMENT

These resaerches are supported by the Délegation Générale pour l'Armement (procurement contract: 04 51 022 00 470 27 75) and the Institut Universitaire de France, we want to thanks. Others thanks to the HUMAINE network of excellence and the european project FEELIX.

Movies of the experiment of fig 5 available on:
http://www.etis.ensea.fr/~neurocyber/giovannangeli/home.htm

REFERENCES

[1] J. Gibson, *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.

[2] F. Varela, E. Thompson, and E. Rosch, *The Embodied Mind*. MIT Press, 1991.

[3] V. Klingspor, J. Demiris, and M. Kaiser, "Human-robot-communication and machine learning," *Applied Artificial Intelligence Journal*, vol. 11, pp. 719–746, 1997.

[4] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," 2002.

[5] M. Nicolescu and M. Mataric, "Learning and interacting in human-robot domains," *Special Issue of IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 31, no. 5, pp. 419–430, 2001.

[6] T. W. Fong, C. Thorpe, and C. Baur, "Robot, asker of questions," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 235–243, 2003.

[7] P. Gaussier and S. Zrehen, "Perac: A neural architecture to control artificial animals," *Robotics and Autonomous System*, vol. 16, no. 2-4, pp. 291–320, December 1995.

[8] P. Gaussier, C. Joulain, J. Banquet, S. Leprêtre, and A. Revel, "The visual homing problem: an example of robotics/biology cross fertilization," *Robotics and autonomous system*, vol. 30, pp. 155–180, 2000.

[9] C. Giovannangeli, P. Gaussier, and G. Dilles, "Robust mapless outdoor vision-based navigation," in *Proc. of the 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2006)*. Beijing, China: IEEE, 2006.

[10] P. Andry, P. Gaussier, J. Nadel, and B. Hirsbrunner, "Learning invariant sensorimotor behaviors: A developmental approach to imitation mechanisms," *Adaptive behavior*, vol. 12, no. 2, pp. 117–138, october 2004.

[11] C. Giovannangeli, P. Gaussier, and J.-P. Banquet, "Robustness of visual place cells in dynamic indoor and outdoor environment," *International Journal of Advanced Robotic Systems*, vol. 3, no. 2, pp. 115–124, jun 2006.

[12] P. Gaussier, A. Revel, J. Banquet, and V. Babeau, "From view cells and place cells to cognitive map learning: processing stages of the hippocampal system," *Biological Cybernetics*, vol. 86, pp. 15–28, 2002.

[13] J. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness," *Behavioral and Brain Sciences*, vol. 24, no. 5, pp. 939–1011, 2001.

[14] C. Watkins, "Learning with delayed rewards," Ph.D. dissertation, University of Cambridge, 1989.

[15] B. Widrow and M. E. Hoff, "Adaptive switching circuits," *IRE WESCON Convention Record*, vol. 4, pp. 96–104, 1960.

[16] F. Kaplan and P.-Y. Oudeyer, "Maximizing learning progress: an internal reward system for development." in *Iida, F. and Pfeifer, R. and Steels, L. and Kuniyoshi, Y., eds.* Springer-Verlag, 2004.

[17] P.-Y. Oudeyer, "Intelligent adaptive curiosity: a source of self-development." in *Proc. of the Fourth Int. Workshop on Epigenetic Robotics: Modelling Cognitive Development in Robotics Systems*, L. Berthouze, H. Kozima, C. Prince, G. Sandini, G. Stojanov, G. Metta, and C. B. Eds, Eds., vol. 117, 2004, pp. 127–130.