Interactive teaching for vision-based mobile robots: a sensory-motor approach

Christophe Giovannangeli and Philippe Gaussier

***Abstract*—For the last decade, we have developed a vision-based architecture for mobile robot navigation. Our bio-inspired model of the navigation has proved to achieve sensory-motor tasks in real time in unknown indoor as well as outdoor environments. We address here the problem of the autonomous incremental learning of a sensory-motor task, demonstrated by an operator guiding the robot. The proposed system enables a semi-supervision of the task learning and is able to adapt the environmental partitioning to the complexity of the desired behavior. A real dialogue based on actions emerges from the interactive teaching. The interaction leads the robot to autonomously build a precise sensory-motor dynamics that approximates the behavior of the teacher. The usability of the system is highlighted by experiments on real robots, in both indoor and outdoor environments. Accuracy measures are also proposed in order to evaluate the learned behavior as compared to the expected behavioral attractor. These measures are used first in a real experiment and then in a simulated experiment enabling to point out the interest of a real interaction between the teacher and the robot.**

***Index Terms*—Mobile robots, Navigation, Robot vision systems, Intelligent robots, Learning systems, Cooperative systems.**

I. Introduction

TASK specification in autonomous robotics has received an increasing interest. It is today admitted that autonomous mobile robots should be designed with a minimal prior knowledge on the tasks to perform so that the robot can adapt to unpredictable situations characterizing the dynamical nature of real environments. The robots should also constitute their skills via interactions with their physical and social environment where they build up experiences from their the sensory-motor interactions [1] leading their own cognition to enact a subjective world [2], also called the *Umwelt* [3]. In this context, Human-Robot Interactions (HRI) are thought to be a very efficient means to specify some various tasks to a robot [4] and to catalyze its sensory-motor learning [5]. HRI are moreover a key point for designing operational or social and interactive robots [6], [7]. This paper investigates the use of HRI for the learning of navigation tasks.

In [8], several problems linked to the autonomous localization and mapping of an environment are pointed out. We summarize here the main points relevant in our approach: 1) The nature of the "noise" on the physical measurements is generally context-dependent. For example, a non-calibrated panoramic camera can induce a biased error of the landmarks position measurement, depending of the robot orientation [9]. Whereas several statistical methods can deal with centered noises, it is much more difficult to detect a conditionally biased noise, characterized by a mean value and a standard deviation which depend on unexpected (hidden) dimensions of the robot state. 2) The required memory (and consequently the required computation time) increases with the size of the environment and the complexity of the internal representations. In the future, it will be crucial to give a bound to the size of the internal representations in order to develop real-time robotics architecture for pseudo-infinite environments 3) The algorithms have to deal with the correspondence problem (or data association problem) to reliably determine if two sensorial measurements taken at different time steps correspond to the same physical point in the environment or not [10]. For a long time, this problem has been treated as stochastic, leading the community to develop algorithms trying to reveal the hidden Markovian model of the environment. Yet, psychology has early identified the ambiguous nature of the perception (Gestalt theory): the so-called multi-stability of the perception implies that a perfectly well defined sensory stimulus can have two antagonist interpretations (Necker's cubes, Rubin's figure and other artistic creation are good examples). This ambiguous nature of the perception should question roboticians whether such a hidden Markovian model of the environment does really exist or whether it is meaningless to try to remove sensorial ambiguities. 4) The dynamical nature of the environment induces environmental changes. During the robot lifelong, localization cues used by the robot may disappear. While the environment is changing, the functioning domain of classical algorithms shrinks until the system no longer works. Hence, a crucial issue in the future will be to provide our robots with re-learning strategies, enabling the adaptation of their knowledge to the environmental changes before their behaviors becomes totally unrelevant. And finally, 5) the robot is confronted to an action selection problem during the building of its internal representations. Robots will have to be endowed with planning strategies in order to select interesting actions in a partially known environment. The meta-learning theory, for instance, has early claimed that a smart selection of the prototypes for the learning can increase the development speed of the robot [11], [12]. Recent works insist on the fact that selecting the action that maximizes the learning progress makes the robot becoming curious and enables it to develop faster [13], [14]. [15] also studied the complexity of greedy mapping algorithms in deterministic environments. Moreover, developers are confronted to a trade-off between rapidity and reliability of the system: accuracy of

The authors are with the Neurobycernetic Team of the ETIS Lab, CNRS UMR 8051, Cergy-Pontoise University, ENSEA, 2 Avenue Adolphe Chauvin, 95302, Cergy-Pontoise Cedex, France.

Corresponding author: christophe.giovannangeli@gmail.com

Manuscript received January 24, 2008; revised July 1, 2008.

fast incremental algorithms strongly depends on the quality of the sensory measurements, whereas statistical methods, asymptotically more accurate, require numerous examples of the environment to build a consistent map. In spite of all this limitation, SLAM methods exhibits impressive performance in long term navigation, when coupled with visual recognition system, in order to deal with the correspondence problem and the environmental changes [16]. Nowadays, it is possible to map large indoor environments, with monocular, stereoscopic, or catadioptric vision systems, though the scaling factor for larger and less controlled environments raises some rarely addressed questions: internal representation size explosion, fusion of isolated map, unreliability of the wheels odometer on rough terrain, ground planarity hypothesis. The sensory-motor teaching for mobile robot navigation, this paper focuses on, has also demonstrated to be very efficient as regards to some of these drawbacks [17], [18].

In the following, we will propose a visual navigation architecture bootstrapped for task specification by imitation that can be useful in many domains in which patrolling or exploring missions are considered. This system will be shown to enable a naive human operator to intuitively teach an autonomous robot to follow a visual path or to perform a homing task. The teacher guides the robot in a task like a visual path following or a homing task, and the robot has to reproduce it. The guidance of the robot by means of a joystick will be used as a simplification of a process of imitation (other works in our lab focus on this aspect [19], [20]). In [21], the problem of task specification is treated as the estimation of a sequence of concurrent behaviors already mastered by the robot (which are likely to have been acquired during the learning phase). The authors also point out that *acting* can provide a basis for a non-verbal human-robot communication and appears as a smart way for the robot to exhibit that it requires some help from the teacher. The idea that the robot could ask questions to its teacher has already been evaluated for example in the collaborative control of [22]. *The robot asks 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 into the symbolic language the robot understands. As a general rule, task specification is performed at a very high symbolic level under the dictatorship of the teacher. However, most of the robotic architectures dedicated to imitation need to separate the learning phases and the performance phases. Yet, lifelong learning constraints [23] imply that the robot must be able to learn while currently freely evolving in the world. A less unilateral process for task specification could emerge from an interactive process of training in which the teacher corrects the robot while the robot tries to imitate the teacher, as proposed in this paper.

Imitation has already proved its interest in machine learning and more specifically in robot skill learning, as illustrated by various studies since the last fifteen years [24], [25], [26], [27], [28], [29], [30], [31]. Theoretical studies have also been undertaken, as in [32], which presents a general formalism for performance metrics on humanoid imitation tasks and illustrates the need of a general framework in order to evaluate the relative accuracy of different algorithms. However, the imitation as a real dynamical and continuous HRI has rarely been stressed (rare examples are [19], [30], [20]). Most of the imitation learning and teaching methods are composed of a phase of demonstration (the learning) and a phase of performing (the reproduction of the knowledge) but rarely the imitation has been treated as a real dialogue based on a language of actions between the robot and the teacher, alternating between imitation and performance phases. In [33], [34], [35], for example, a demonstrator tries to teach a humanoid robot to grasp an object. The study compares an imitation strategy based on the recording of the joint positions of a human and an embodied demonstration based on the recording of the joint position of the robot while the teacher physically moves the robot arm in order to demonstrate the task. The authors points out that the teacher demonstration does not take into account the embodiment of the robot whereas the realization of the task by the robot during the learning is far more pertinent. Although they insist on the role of the observation of the performed task to help the teacher to understand the robot skill and to prepare the following demonstration, they do not take benefit of the intervention of the teacher during the task realization. Indeed, in the context of the interactive teaching, learning and demonstration phases ought to be gathered in order to provide a rich and natural communication which could improve the development of the robot skills: by imitating a teacher, the robot could experiment the behavior that has to be learned. By *acting and reacting* to the teacher orders, the robot should freely exhibit its mastery of the task while in parallel improving its learning [5]. At the same time, the observation of the robot behavior enables the teacher to see and intuitively measure the effect of his teaching and can help him to discover how to efficiently correct the robot. Although this procedure appears as a non-verbal, non-symbolic communication, we claim it is nevertheless a very rich communication [19] able to catalyze the learning of the robot. In such an interactive context, a strong autonomy of decision as well as a strong autonomy of the learning is necessary. As humans are involved, rapidity, precision and adaptation of the learning are also required.

This paper first presents our robots and its visual system enabling to create a continuous state space. Then, we will propose a bootstrap[1] for the PerAc architecture [36] that enables the semi-supervised learning of a sensory-motor behavior (a visual path, a homing behavior). The couple architecture-equations enables to adapt the partitioning of the environment to the complexity of the task. The system does not separate learning and performing phases, which are scattered in time according to the rhythm of the interaction. The system will be evaluated in a real indoor environment by means of accuracy measures between the performed trajectory and the expected behavioral attractor of the robot dynamics. The interest of the interaction between the human and the teacher during the learning, especially the importance to adopt a proscriptive teaching strategy allowing the robot to commit its own errors,

[1] a parallel and supplementary architecture emcompassing a first architecture to control its learning dynamics.

will be experimentally illustrated using the proposed measures.

II. Methods and Materials

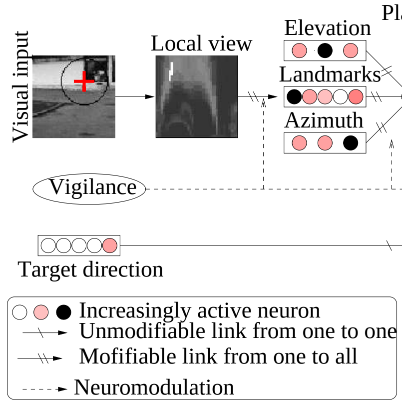


Fig. 1. Block diagram of the architecture. Our architecture for place recognition is composed of a visual system that focuses on points of interest and extracts small images in log-polar coordinates (called local views), recognized as landmarks (see fig. 2). Next, a merging layer compresses the *what* and *where* information, to allow place recognition. By incorporating our visual place recognition system in a PerAc architecture, it is possible to create an attractive behavior to the goal. Each new learned place is associated with a movement which is triggered when the robot recognizes the place. The vigilance signal triggers a wave of one shot learning of the landmarks related to the current location, next of the current place code, in order to allow the learning of the current place-action association.

Among the various methods to create spatial behaviors, the PerAc (Perception-Action) architecture [36] has demonstrated to be particularly adapted for online sensory-motor learning. A PerAc architecture may underlie many various skills in mobile robotics: guidance [37], local navigation in indoor [17] and outdoor environments [18], planning [38], reproduction of a temporal sequence of actions [27], as well as in the control of actuators with multiple degrees of freedom: arm robot control [30], [39], gaze direction control. This architecture is able to learn online sensory-motor associations. In this paper, the PerAc architecture is coupled with a bio-inspired model of visual place-cells computing a robust localization gradient in indoor as well as in outdoor environments [40], in order to perform local navigation tasks [17], [18].

Fig. 1 summarizes the visual processing chain for the place recognition. A place is defined by a spatial constellation of online learned visual features (here a set of triplets *landmark-azimuth-elevation*) compressed into a place-code. The constellation results from the merging a *what* information and a *where* information provided by the visual system that extracts local-views in log-polar coordinates, centered on points of interest. Fig 2 illustrates the autonomous landmark extraction mechanism.

A remarkable property lies in the built-in generalization capability of the system (see [40] for more details). To summarize, a place-cell encoded in location A responds maximally in A, and creates a large decreasing place-field around A. In the experiment of the fig. 3, the robot learns 5×5 positions regularly located in a classical working room (fig. 3.a). The fig. 3.b shows the created place-fields for each place-cell in the whole environment, corresponding to a localization gradient.

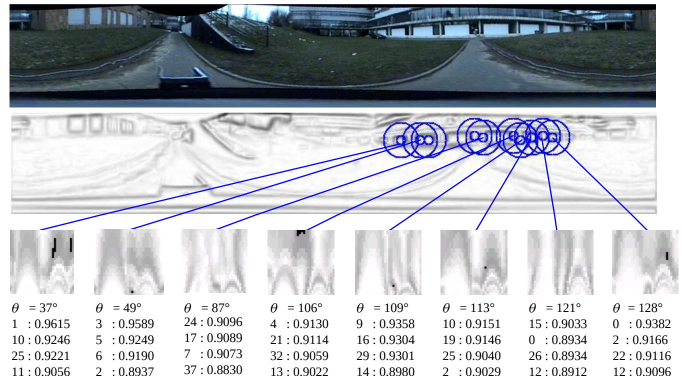


Fig. 2. Illustration of the landmark extraction mechanism: the gradient of a panoramic image is convolved with a DoG filter. The local maxima of the filtered image correspond to points of interest (centre of the circles). Here, the eight first focus points are displayed. The system focuses on these points to extract local views in log-polar coordinates corresponding to landmarks. The system also provides the bearing of the focus points by means of a magnetic compass. For each extracted local view, the identity of the four most recognized landmarks and their recognition levels are given.

In a first approximation, a place-cell activity can be estimated by a noisy Gaussian curve:

$$p_{x_l}(t) \simeq e^{-\frac{\|x_l - x(t)\|^2}{\sigma^2}} + \epsilon^P(t).$$

with $p_{x_l}(t)$ is the activity in $x(t)$ of the place-cell encoded in x_l, with σ expresses the extent of the place-field which is linked to the distance of the landmarks, and with $\epsilon^P(t)$ a noise induced by the uncertainty of the azimuth measurements, the camera discretization, and the dynamical nature of the environment.

The learning of several locations creates overlapping place-fields and also leads to the paving of the space when the learning of new locations is triggered by the detection of low place-cell activities (according to a given threshold). A mathematical consequence of the *what* and *where* merging is the following: the shape of the place-field is homothetic with the shape of the environment [41], [40] (*i.e* the place-fields extend with the distance to the landmarks). As regard to the problem of the size of the world representation, our system exhibits a real interest. The system builds its own metrics based on the azimuthal shifts of the landmarks and their recognition level. Hence, the dimensionality of the internal representation is not given by the Cartesian size of the explored area but rather by its visual regularity (*i.e.*: if the distance to the landmarks were infinite, the world description would be reduced to a single PC) [40]. The computational load and the memory requirements has been proved to be a linear function of the number of learned landmarks [9]. Hence, the learning of a loop in a large outdoor environment uses the same computation load and memory requirement as the learning of a loop in a smaller indoor environment (see experiments of section IV). To our knowledge, extremely few algorithms exhibit such a property. Moreover, neither Cartesian nor topological map building is required for the localization, since the world acts as an outside memory [42]. As long as the learned features of a location persist in its neighborhood, the robot is able to self-localize without map building.

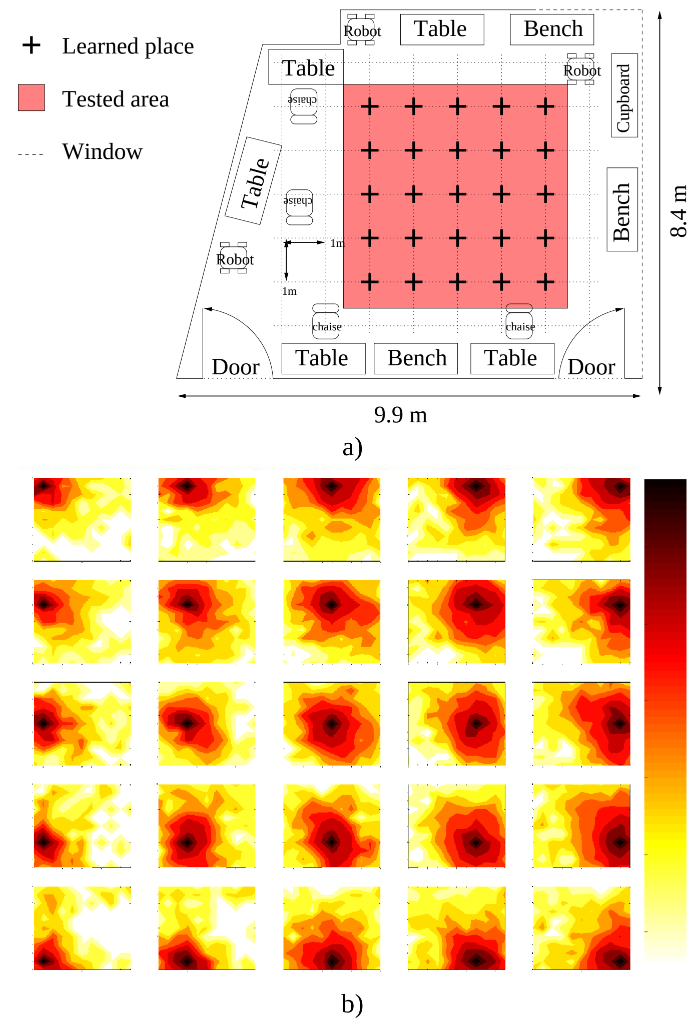


Fig. 3. a) Working room used in the experiment of fig. 3.b. 25 places are regularly learned in the room and tested in the whole room. b) Activity of 5×5 place-cells regularly encoded in the working room of the fig. 3.a. A competition between all the place-cells leads to the paving of the environment.

The problem we address in this paper concerns a more general class of algorithms, based on place recognition, which can lead to an adaptive environmental paving. For example, GPS measurements, a triangulation system via external references, classical SLAM, vision-based SLAM or topological approaches provide the information to use the methods we will present. An intuitive approach to achieve visual navigation using a localization gradient could be to use a hill-climbing algorithm on the place recognition level of a goal cell (a particular PC). Unfortunately, even if the robot could keep a direction as long as the recognition level increases, a strong initialization problem occurs each time a new action has to be chosen. The noise on the place recognition level can also induce local maxima. The duration of each movement represents a critical parameter for the convergence of such an algorithm. Minimization parallax between a learned place and the current location, inspired by models of the insects navigation [43], [44], could be used to avoid pure 1D hill-climbing methods. As actions are directly computed rather than being learned (thought it is possible [45]), the behavior is not adaptive and the trajectories are stereotyped. Moreover the learning of a trajectory requires massive efforts either on the problem of learning a sequence of actions and detecting place reaching (also called milestone points in [46]) are reached, or on the problem of the cognitive mapping of the learned locations [47]. Finally, the question of the robustness evaluation has rarely been raised. Nevertheless, recent studies [48], [49] propose an improved version of the ALV algorithm [44] able to keep a constant performance level independently of the size of the environment. Several of these limitations can be overcome by using a PerAc Architecture: a simple associative learning between places and actions is able to create a sensory-motor attraction basin, for homing or path following behaviors (see fig.1 for the architecture). The problem of building a policy of actions has often been stressed in the literature of reinforcement learning [50], [51], [52], [53], [54], [55] but we claims that the PerAc architecture is extremely efficient for spatial behavior learning since it embeds the problem of the environmental partitioning as well as the problem of action policies learning[2]. The next section will address the problem of the autonomous building of behavioral attraction basins by human-robot interactions. The problem is treated as a machine learning problem through an interactive demonstration.

The various platforms and electronic equipments we used to study mobile robot navigation are the following (see fig. 4.a):

- Koala K-Team, pan-tilt camera, magnetic compass.
- Koala K-Team, omni-directional camera, magnetic compass.
- Pioneer 2 AT ActivMedia, stabilized platform, pan-tilt camera, magnetic compass.

For outdoor experiments on rough terrains, we built a gyro-stabilization platform in order to deal with the effects on a non-planar ground (see fig. 4.b).

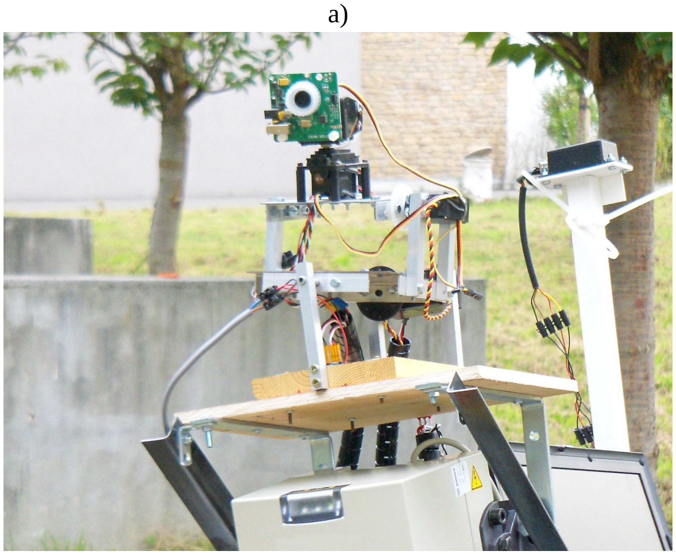
III. Learning and refinement of a spatial behavior: a sensory-motor approach

The presented work proposes a reformulation of the problem of autonomous spatial behavior learning already addressed by the various reinforcement learning methods [58], [59], such as Q-Learning [52], [53], TD(λ) [51], Policy Gradient Reinforcement Learning (PGRL) [54], or Value and Policy Search (VAPs) [55] ... Our approach differs from them because the continuity of the state and action spaces is not a particular context in which the algorithm has to be extended but a basic assumption that has been guiding the design of our architecture. Our approach also differs because we aim the design of a complete architecture (able to control real robots) rather than a theoretical algorithm isolated from its architectural layout. Moreover, classical reinforcement learning algorithms try to affect a score to each encountered state or state-action unit of the environment corresponding to an expected reward. Based on the propagation of the reward, reinforcement algorithms are too slow in convergence in a continuous environment because

[2]we prefer in our school of thinking the terms behavioral dynamic instead of action policy, referring more to the psychological literature on learning and control of human coordination and perception.

they first need to partition (adaptively or not) the environment before the reinforcement learning algorithm can perform. Methods for the partitioning of the whole state-action space has also been proposed [60]. As human-robot interactions are concerned, we can not accept a slow acquisition of the behavior (even if sure and optimal), the acquisition (and the usability of the knowledge) must be performed in a very short time. If an algorithm is allowed to spend time to estimate the state space, this time should be used in parallel to estimate the topology of the environment. The estimation of the state space topology gives access to a cognitive map which can compute a latent learning of many unrewarded paths [61], [62], [63], [64], [65], [41], [66], [38]. Evidences of a such a latent learning have been given in mammalian species since 1948 by Tolman [67], showing that the time for a rat to find a goal does not decrease once the reward is found, but latently decreases with the number of experiences of the future goal path before the discovery of the reward. Thus, once a reinforcement occurs in a given state, efficient (but sub-optimal) strategies are directly available from each visited places.

a)



b)

Fig. 4. a) Wheeled and legged robots used to study bio-inspired navigation. The left robot used an omni-directional camera, the right robot uses a firewire camera mounted on a gyro-stabilized pan-tilt platform, the wheeled robot in the centre uses a classical pan-tilt camera. All the robots are provided with a magnetic compass (CMPS03). However, in [56], [57], we shown that the magnetic compass can be replaced by a visual compass associated to a path integration system. We also tried to adapt the system to legged robot like Aibo. b) Gyro-stabilization platform used for experiment on rough terrains.

Moreover, continuous state and action spaces are generally treated as discrete after quantization. What has been encouraging researches in reinforcement learning is the proof of optimality which already exists for various algorithms, mostly in discrete and non-stochastic state and action spaces [52], [68], [69]. However, convergence towards optimal solutions in stochastic and continuous spaces is not guaranteed for most of the reinforcement learning methods. Q-Learning for example is proved to converge only locally for a certain class of problems that has continuous state and action spaces [70]. It has also been highlighted that reinforcement learning algorithms may diverge when a function approximation is used instead of a look-up table [71]. On the contrary, our sensory-motor architecture takes into account both continuousness of both the state and the action spaces. This paper will show that a continuous action space enables the measure of an error helping for the adaptive partitioning of the continuous state space. Moreover, since the suboptimal solutions found by the Nature for the animal navigation are more robust then the current engineering solutions, we can wonder about the need of an optimal algorithm for the learning of spatial behaviors. We can also wonder about the interest of convergence proofs as compared to the time to obtain an efficient sub-optimal behavior (as regard to an external measure). Works like [72] has highlighted that reinforcement learning algorithms can really perform better when initialized with a sub-optimal policy. The sub-optimal solutions computed by our architecture could be used to initialize reinforcement learning algorithms.

Actually, some limitations of classical reinforcement learning algorithms can be overcome by bootstrapping a PerAc architecture (see fig. 1 [36]). Each PC is associated with a movement to trigger when the corresponding place is recognized. If the PCs and the actions are defined in the frame of a competitive structure, a minimum of three place-action associations around a goal creates a behavioral attractor, leading the robot trajectories to converge towards the goal from each place in the attraction basin. Learning is equivalent to shape this basin in order to create an accurate behavioral attractor. Homing or route following behaviors (see fig. 5 and 6) can be learned in one shot. Even though human assistance could speed-up the convergence [72], classical reinforcement learning methods are not efficient with so few learning samples.

A. HRI and the PerAc architecture

We investigate here how the PerAc architecture can underlie the learning of navigation tasks in the frame of an intuitive human-robot interaction. In our previous experiments of visual homing or path following (see figs. 5 and 6), the learning was totally supervised by a human who positioned the robot in a precise location with a precise orientation, or was generated by an ad-hoc process (moving around a goal position to learn it from different positions). Yet, the PerAc architecture is particularly well designed for the real time online learning

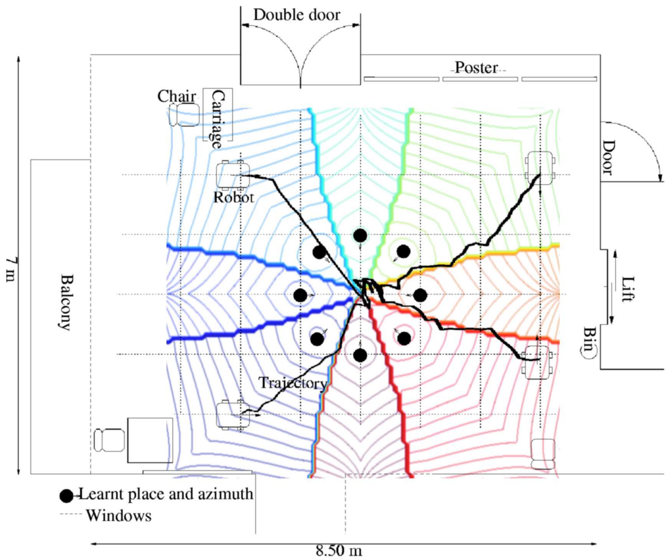


Fig. 5. Real trajectories of homing in an indoor environment with an omni-directional camera. 8 places (black circles) are learned at 1 m from the goal (size of the square on the floor). The theoretical place-fields are superposed with the map and the trajectories.

Fig. 6. Outdoor environment and looped sensory-motor trajectory. Arrows represent the learned positions and the associated movements. The robot closes the loop of about 100 m in 20 mn. The system is slow because the whole architecture was executed by a single sequential program (September, 2005).

of skills in that sense that its goal is to learn associations that occur through direct volontary experience (concept of enaction [73]). Hence, guiding the robot through the task should be enough and more ergonomic in order to specify the task to the robot than an explicit symbolic communication as used in [22] or [74].

In the context of lifelong learning [23], we are presently interested in addressing the problem of the semi-supervised building of a behavioral dynamics and its refinement. Besides, we focus here on the capability of the robot to autonomously learn a sensory-motor task by interacting with a human. Being guided by the human, the robot learns places and is able to merge the action associated with the current state (here places) to the action imposed by the teacher. We use a joystick to guide the robot in the same way as a dog could be guided with a leash, but a visual tracking of the teacher could also have been possible (really close to an imitation process). We propose here an autonomous architecture enabling the robot to learn in one shot a new place-action association and to adapt the movement associated to the previous place according to the sensory-motor error generated during the crossing of this place.

In the PerAc architecture, two learning stages can be controlled: the sensory learning (environmental partitioning) and the sensory-motor learning (policy of action learning). In classical task specification in unknown environment, the environmental partitioning has to be stabilized before the navigation can be performed. Here, we save time by the simultaneous one shot learning of both the sensory state space and the sensory-motor associations. Each time a sensory state is learned, a motor action is instantaneously associated with it. A vigilance signal will be responsible for triggering this wave of learning (see fig. 1).

B. Movement adaptation

We consider two binarized signals for the bootstrap of the sensory-motor learning. The first signal is the vigilance signal $V(t)$ which triggers the waves of one-shot learning. The second signal $\epsilon(t)$ corresponds to a learning rate. It is used as a modulation for both the one-shot learning and the adaptation. The neural architecture is given in fig. 7. In our architecture, $\epsilon(t)$ spikes each time a place transition occurs (hence also each time the vigilance signal spikes). The group of neurons A^P (which elements are a_k^P), performing the motor learning, is inspired from the Widrow-Hoff (WH) learning rule [75] but other rules are possible[3]. The main difference with a classical WH learning rule is that our rule is composed of two terms. A term performing a one shot learning computed as the classical gradient of a WH learning rule and a term computed according to the previous gradient computation, corresponding to a delayed learning rule.

In the following, the activity of the place-cells is binarized: $p_i^+(t)$ 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. The signal $\epsilon(t)$ corresponds to a place transition ($\epsilon(t) = 1$ when a place transition occurs and $\epsilon(t) = 0$ otherwise). It can be defined as: $\epsilon(t) = \sum_{i=1}^{n_P} \left[p_i^+(t) - p_i^+(t-dt)\right]^+$, with n_P the number of place-cells, and $[x]^+ = x$ if $x > 0$.

The actions are defined by population of neurons: each neuron k in an action group corresponds to a particular orientation $\frac{2.k.\pi}{n_A}$, n_A being the number of neurons coding an action ($n_A = 61$ in our architecture). The activity of the group $A^R(t)$, providing the performed movement between $t-dt$ and t in the direction $\theta(t)$, is a Gaussian curve, centered on the neuron corresponding to the orientation $\theta(t)$. Hence:

$$a_k^R(t) = e^{-\frac{|\Delta_k^\theta(t)|^2}{\sigma}} \quad (1)$$

[3]The Hebbian learning rules has been rejected because the time to learn a new action would have been greater or equal to the whole time of learning (the longer the system has already learned, the longer learning something else will be). Moreover, the Hebbian learning rule needs to be shunted by means of a multiplicative term $1 - \omega_{ik}$ so that the weight could be in $[0, 1]$ (corresponding to a Grossberg rule), creating a dynamics very close to the WH learning rule.

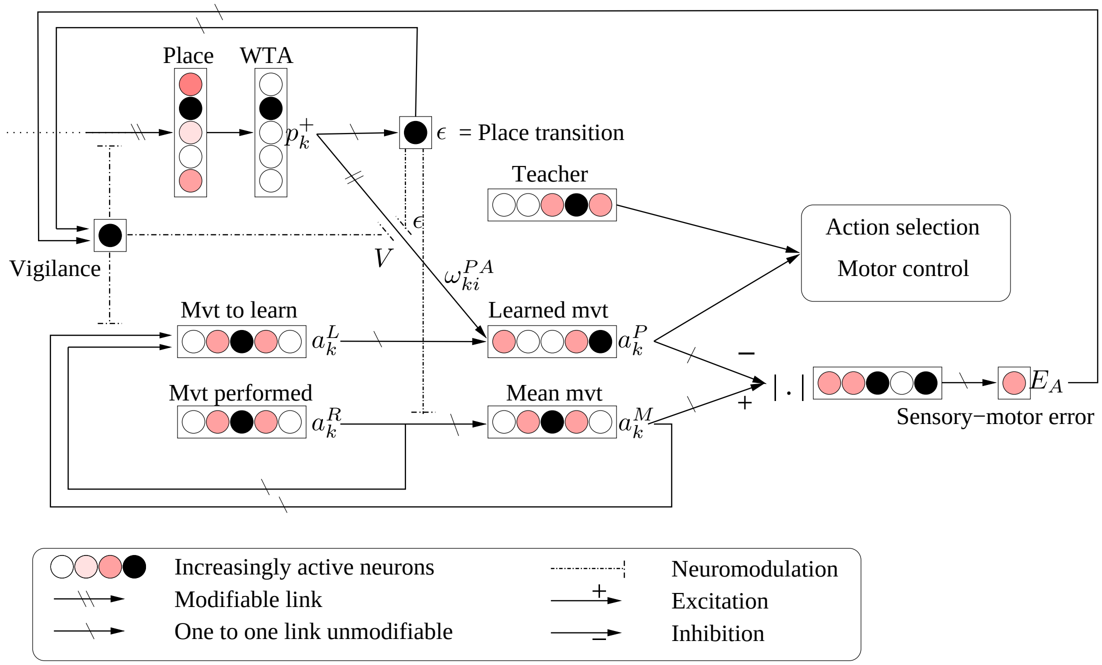


Fig. 7. Modified PerAc Architecture enabling either the one-shot learning of places and place-action associations or the refinement of the sensory-motor dynamics. The computation of a signed angular error between the mean performed movement and the predicted movement in a given place enables to adapt the movement associated to this place. The one-shot learning of the landmarks, the constellations, the places and the place-action associations is triggered by a vigilance signal, whereas the adaptation is performed continuously, each time a place transition occurs.

with $\Delta_k^\theta(t) \in]-\pi, \pi]$ the shift between the favorite direction $\frac{2.k.\pi}{n_A}$ of the neuron k and the performed movement $\theta(t)$ (here, $\sigma = \frac{\pi}{6}$).

The neurons of the group A^M provide the mean movement and are defined as:

$$a_k^M(t) = \epsilon^M.a_k^R(t) + [a_k^R(t-dt) - I^R.\epsilon(t)]^+ \qquad (2)$$

with ϵ^M a rate avoiding $a_k^M(t)$ to be greater than 1 until $\frac{1}{\epsilon}$ steps without reset ($\epsilon^M = 0.001$ for example), with I^R a strong positive signal resetting the memory of a_k^M ($I^R = 1000$ for example).

The activity of the kth input neuron for the motor learning $a_k^L(t)$ (output to learn) is computed as follow:

$$a_k^L(t) = a_k^R(t-dt).V(t) + \frac{1}{a_{max}^M(t)} a_k^M(t-dt).\epsilon(t).(1-V(t)) \qquad (3)$$

with $a_{max}^M(t) = \max_{k=1..n_A}(a_k^M(t))$, used for the normalization (a_k^R being already normalized). $a_k^L(t)$ 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 the delayed adaptation). The mean movement is reset by the $\epsilon(t)$ signal (see fig. 7), each time a place transition occurs.

The equation for updating the activity of the neurons a_k^P performing the sensory-motor learning is the following:

$$s_k(t) = \sum_{i=1}^{n_P} \omega_{ik}^{PA}(t) p_i^+(t) \qquad (4)$$

$$a_k^P(t) = V(t).a_k^L(t) + (1 - V(t)).\left(\frac{s_k(t)}{s_{max}(t)}\right) \qquad (5)$$

In this equation, $s_k(t)$ is the predicted activity of the kth neuron of the group. ω_{ik}^{PA} is the weight of the connection between the ith place-cell and the kth action neuron. Finally, $s_{max} = \max_{k=1..n_A}(s_k)$ is used for the output normalization. More precisely, $a_k^L(t)$ is the desired output (the future action to predict, explicitly given by the input group A^L called *Mvt to learn* in fig. 7.). The equation 4 corresponds to the predicted output and the equation 5 provides the effective output computed either as the normalized prediction or as the desired output (which is also normalized) during a one-shot learning cycle (no prediction being available before the one-shot learning). Most of the signals (inputs and outputs) are normalized in order to compute the sensory-motor error E_a, defined as the difference between the performed movement and the learned movement for a given place: $E_a(t) = \sum_{i=1}^{n_A} |a_i^M(t) - a_i^P(t)|$.

The update of the synaptic weights is performed after the update of the neurons activity according to the following equations:

$$\frac{d\omega_{ik}^{PA}}{dt} = (G_{ik}^i(t) + G_{ik}^d(t-dt)).\epsilon(t) \qquad (6)$$

with:

$$G_{ik}^i(t) = (a_k^L(t) - s_k(t)).p_i^+(t).V(t) \qquad (7)$$

$$G_{ik}^d(t) = (a_k^L(t) - s_k(t)).p_i^+(t).(1-\epsilon(t)) \qquad (8)$$

In this equation, two gradient terms are computed: G_{ik}^i (instantaneous gradient) which is the classical WH gradient with a term of vigilance modulating the learning and G_{ik}^d (delayed gradient) which computes a gradient if no learning or adaptation occurs. During a one-shot learning cycle (when $V(t)$ and $\epsilon(t)$ spike), the new place is associated with the current action by means of the not null terms $G^i_{ik}(t)$ in the equation 6. Otherwise, a delayed adaptation is performed each time $\epsilon(t)$ spikes by means of the term $G^d_{ik}(t-dt)$ (the previous gradient). Hence, the adaptation of the movement in a place is performed only once the robot has left the place and will only be available 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 and the contents of the learning.

The remaining question concerns the control of the vigilance signal: Which are the important signals for the autonomous partitioning of the environment corresponding to a refinement at the sensory level?

C. Adaptive partitioning of the environment

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. The simplest solution to trigger the coding of a new place is to fix a low threshold t^-_P on the place-cell activity. If the activity $p^M(t)$ of the most activated place-cell is under this threshold ($p^M(t) = \underset{k=1..n_P}{max}(p_k(t))$), a new place is learned: $V(t) = \Gamma_0(t^-_P - p^M)$, with Γ_x the Heaviside function: $\Gamma_x(y) = 1$ if $y > x$ and 0 otherwise. This will lead to a regular partitioning of the environment. The threshold t^-_P has to be low enough in order to use the generalization capabilities of the place-field and to minimize the number of encoded place-cells. Such a vigilance signal implies that the size of the place-fields (and also the precision of the spatial encoding) is fixed as illustrated by figs. 8.a and 8.b. Since the sampling of the partitioning controls the precision of the behavioral dynamics, the partitioning of the environment should not be regular but adapted to the desired precision and to the complexity of the trajectory (see fig. 9). For instance, more place-action associations should be encoded during a sharp bend than during a straight line. The system could use the discrimination capabilities of the place recognition in the complex parts of the trajectory and its generalization capabilities in the easier parts. In a more general context, the assumption that a given sensory-motor function $D : S \rightarrow M$ is better approximated if the discretization factor of the sensory space S evolves as the variation $\frac{\Delta D}{\Delta S}$ of the sensory-motor function remains valid (the compression factor is adapted to the variations of information).

The difference between a regular paving (corresponding to a threshold on the sensory dimension) and an adaptive paving (corresponding to a threshold on the action dimension) is illustrated for a one-dimensional example in the fig. 10. With the regular partitioning, the more the function varies, the higher the error is. Yet, when the function is varying, the probability of generating diverging trajectories is higher.

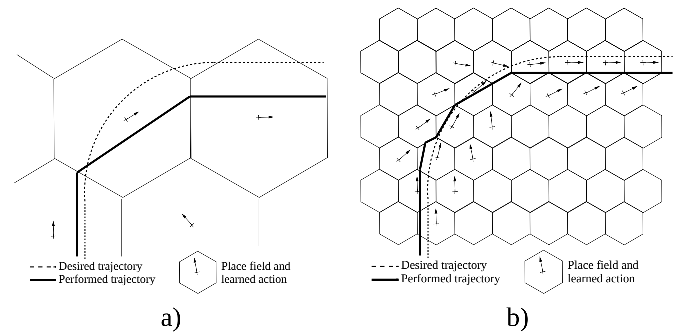


Fig. 8. a) An example of a regular spatial partitioning with large cells. A movement direction is associated to each cell. The precision of the reproduced trajectory depends on the precision of the state space coding (the size of the cells). b) An example of a regular spatial partitioning with small cells. The precision of the reproduced trajectory is higher than in fig. 8.a. However, the cost of the spatial coding in also higher.

On the contrary, with the adaptive partitioning, the density of encoded place-cells increases with the variation of the function to approximate: fewer place-cells are recruited when the function is monotonic and more are used when the function varies. Hence, the more the function varies, the lower the error is.

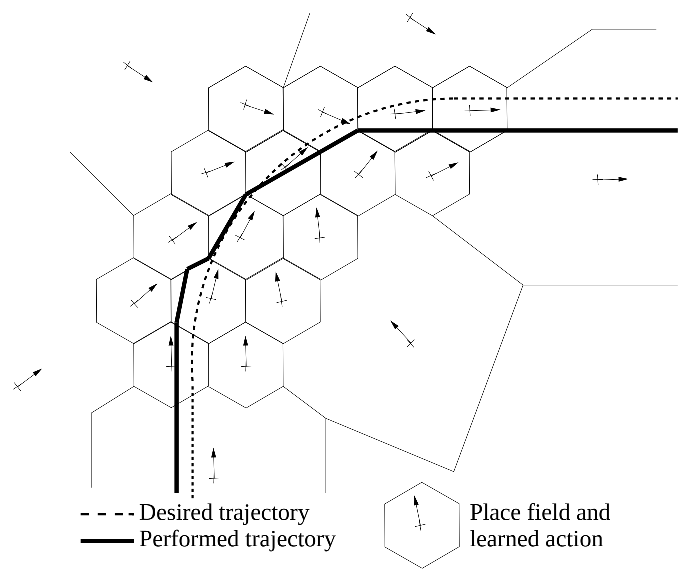


Fig. 9. An example of an adaptive spatial partitioning. The size of the cells is adapted to the complexity of the trajectory. Small cells are used to precisely follow the bend, whereas big cells are used to create a convergence around the trajectory.

The sensory-motor error $E_a(t)$ in each place has been defined as the difference between the predicted and the performed action. It stands for the parameter $\frac{\Delta D}{\Delta S}$. 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. A threshold $t^+_{E_a}$ on the sensory-motor error is responsible for the accuracy of the behavior during a bend. For example, if $t^+_{E_a}$ corresponds to an error of about 30^o, then a 90^o bend should be encoded by at least three place-cells. Thus, this measure can be used to control the vigilance signal in order to adapt the learning location of the place-cells to the complexity of the desired behavior. The vigilance signal is defined as:

$$V(t) = \Gamma_0\left(\left(t_P^{+}-p^M(t)\right).\left([E_a(t)-t_{E_a}^{+}]^{+}+[t_P^{-}-p^M(t)]^{+}\right)\right) \tag{9}$$

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 maximum of the place-cell activities $p^M(t)$ 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.

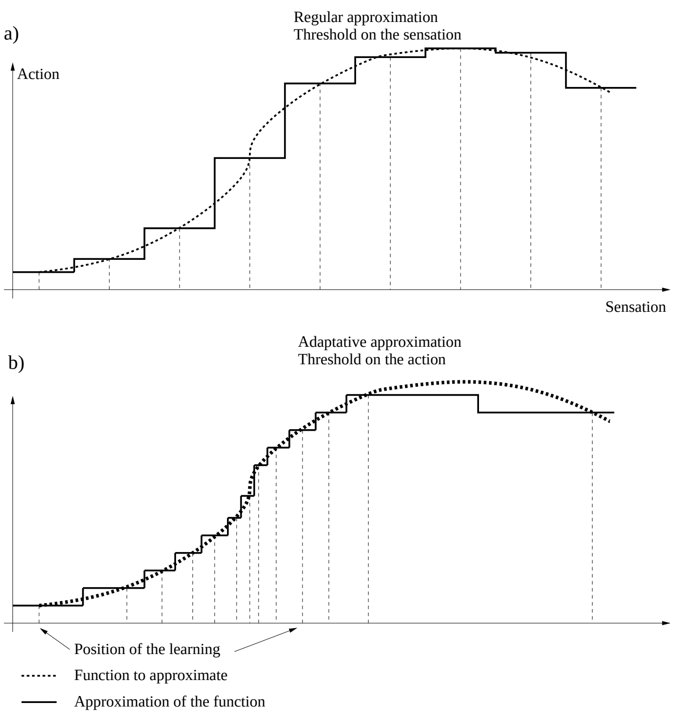


Fig. 10. Illustration on a one-dimensional example of two approximation methods. Fig. a) illustrates the regular partitioning based on a low threshold on the place-cell activity (here corresponding to a given distance between the position of learning on the "sensation" axis) under which a new place is learned. Fig b) illustrates the adaptive partitioning based on a high threshold between the learned action and the action to be learned, over which a new place is learned. The adaptive partitioning is able to reduce the error when the variation of the sensory-motor dynamics (the function to be learned) is high and to create large place-cells in monotonic parts of the sensory-motor dynamics.

D. Simulated environments and simulated teacher

In order to theoretically validate our approach, a simulated environment is used. In this environment, the system creates perfect place-cells since all the possible landmarks as well as their identity and their exact azimuth are provided. In order to simulate the human guidance of the robot, an ordered set D of points d_i that parametrizes the desired trajectory is predefined. A dynamical process, which trajectories converge towards an attractor considered as the optimal trajectory, is then used. The process consists in identifying the closest point d_i in the desired trajectory and in heading for the point $d_{i+\Delta_P}$, with $\Delta_P > 0$ depending on the proximity[4] d of the points d_i ($\Delta_P = 5$ in our simulated environment with $d < 7.5 \times \sqrt{2}$ and with the distance d_r (distance travelled by the agent between each steps) defined so that $d_r < 1$). The dynamical system defining the human guidance is illustrated by the fig. 11.a. The figure 11.b illustrates, for a given set D that parametrizes the desired trajectory, the trajectories generated by the described dynamical system simulating the human guidance. The generated attractor corresponds to the expected robot behavior after the learning (ie: the attractor corresponds to the desired trajectory).

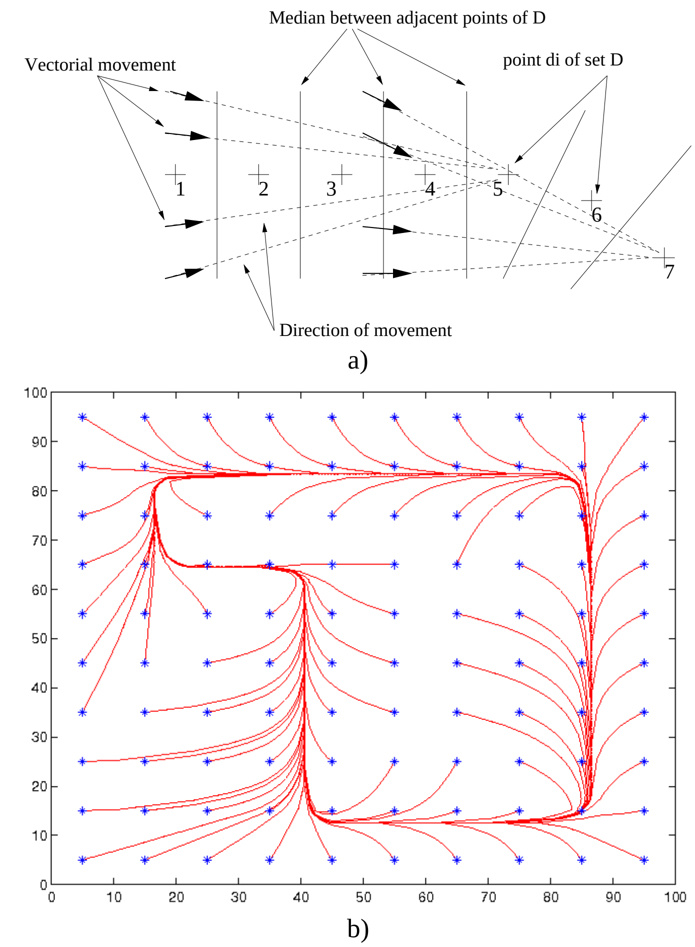


Fig. 11. a) To simulate the human guidance, an ordered set D of points d_i that parametrizes the desired trajectory and a dynamical process which trajectories converge towards an attractor considered as the optimal trajectory is used. b) Trajectories generated by the dynamical process of the fig. 11.a. The trajectories converge towards an attractor defining the optimal trajectory.

E. Validation of the proposed system

The experiment of fig. 12 illustrates the capability of the system to adapt the spatial partitioning to the complexity of the desired trajectory. In this experiment, a human presses a

[4]the unit used to measure distance is the pixel. The size of the following environment is 750×750 pixels.

button when he wants to correct the robot behavior in order to teach him the desired trajectory. The strategy of the teacher (when the button should be pressed?) is the subject of the next section. The figures 12.a and 12.b show the resulting trajectories as well as the attractor of the sensory-motor dynamics of the robot. The attractor is defined as the mean position of the robot for different starting points in the attraction basin after a long time of convergence. The figure 12.c also shows the position of the learned places, superimposed with the attractor. The simulated robot adapts the density of coded locations to the complexity of the desired trajectory: during blends, the robot uses the discrimination capabilities of the place-cells in order to accurately approximate the desired behavior, whereas the system uses the generalization capability of the place-cells in easier parts of the desired trajectory like straight lines.

The use of the sensory-motor error $E_a(t)$ to control the learning of a new location allows to adapt the precision of the spatial partitioning to the complexity of the task. Moreover, precise thresholds do not have to be estimated, but confidence thresholds for the recognition and the non-recognition. The threshold $t^+_{E_a}$ on the sensory-motor error could also be learned.

IV. Human-robot interactions as a cognitive catalyst for the learning of behavioral attractors

In this section, accuracy measures of reproduced trajectories as compared to the optimal trajectory are proposed. The interest of the interaction loop between the human and the robot during the learning, especially the importance of allowing the robot to commit its own errors, is demonstrated using these measures. Finally, we use these measures in a real indoor environment, by means of a vision-based system which corrects the perspective and enables the tracking of the robot position in the Cartesian space. An experiment in outdoor environment is also proposed.

A. Proposition of an accuracy measure of the trajectory reproduction

The reproduction of a trajectory is a problem frequently addressed in mobile robotics. As optimality is not always reached or tracked among the various algorithms, we propose a measure that could help to compare the generated trajectories to an optimal path (the expected behavioral attractor). Since it could be very long to evaluate the complete behavior in the whole environment or to estimate the optimal behavior in each position, we prefer trying to evaluate the precision of the generated trajectory, from its starting point to its end point with respect to a desired trajectory in order to compare the performance of different algorithms.

Evaluating the spatial precision of a trajectory, independently of the temporal precision, is an extremely hard problem. Indeed, comparing trajectories without time aspects is equivalent to compare the sequence of points defining the two trajectories. We propose two measures in order to compare the optimal trajectory $\{x_i(p)/p \in \{1..P\}\}$ with the reproduced trajectory $\{x_r(t)/t \in [t_i..t_f]\}$. The first equation evaluates the mean distance between the robot position and the closest point of the desired trajectory:

$$e_t = \frac{\int_{t=t_i}^{t=t_f} \underset{p=1}{\overset{P}{min}} \; \|x_r(t) - x_i(p)\|.dt}{t_f - t_i} \qquad (10)$$

This measure is not enough since the robot can navigate very close from the desired trajectory during its whole trajectory but stay very far from a given point of the trajectory. For example, if the robot does not move, the measure e_t is constant. Hence, a second equation has to be introduced. It verifies that the robot has travelled close enough to each point of the desired trajectory:

$$e_p = \frac{\sum_{p=1}^{P} \underset{t=t_i}{\overset{tf}{min}} \; \|x_r(t) - x_i(p)\|}{P} \qquad (11)$$

This second equation is also insufficient since the robot can navigate close to each point of the desired trajectory and then escape very far without increasing the measure. However, the jointure of both equations allows to evaluate if each robot position was always close to a point of the desired trajectory and if the robot has been close to each point of the desired trajectory. Hence, a combined measure may also be used, such as $(e_t + e_p)$.

It must be noticed that each measure varies in opposite manner. The first measure e_t is low at the beginning and increases with the error of reproduction, whereas the second measure e_p is high at the beginning and decreases with the accuracy of the reproduction. At the end of a relatively correct reproduction, e_t should have increased to a weak mean value (the robot has never been far from a point of the desired trajectory) and e_p should have shrunk to a weak value (the robot has been close to each point of the desired trajectory).

However, it is still possible to find some trajectories which are well scored but correspond to a wrong reproduction. For example, if the robot reproduces the trajectory in the opposite direction, the score will be the same as in the correct direction. Moreover, oscillating around the ideal trajectory provides the same score as a straight trajectory. An angular term could be useful. The duration or the length of the performed trajectory could give another estimation of the quality of the reproduction. We consider that the robot has to be able to reproduce the trajectory in the correct direction and with few oscillations of the direction before using these measures. In the following, these two measures will be used in a simulated environment and in an indoor environment. However, it is far more difficult to use these measures in larger environments since they require precise measurements of the robot position. In outdoor environments, a differential GPS seems necessary. In indoor environments, systems based on a network of calibrated camera, tracking the robot across several rooms could be possible. However, we did not have access to these technologies for our experiments. Hence, in the following experiment in the outdoor environment, these measures have not been used, because of the difficulty to estimate the precise position of the robot.

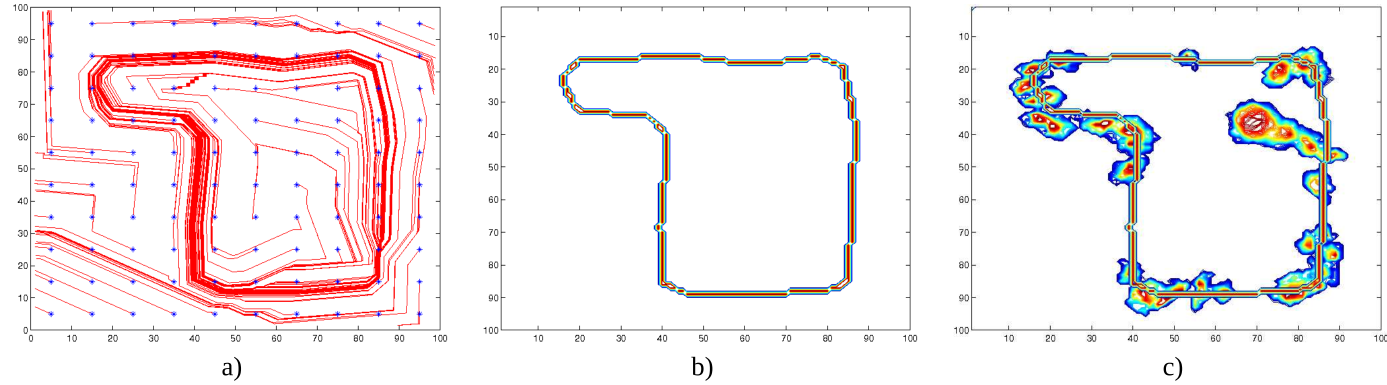


Fig. 12. Training a simulated robot to perform a given trajectory: a human pushes a button to trigger the guidance process defined in fig. 11. a) Trajectories generated after the learning. b) Generated attractor: mean position of the robot for different starting points in the attraction basin, after a long time. c) Position of the learned places, superimposed with the generated attractor. The system has adapted the density of encoded location to the complexity of the desired trajectory. More locations are learned in blends than in straight lines. The system uses the discrimination capabilities of the place-cells in complex parts of the desired trajectory and the generalization capability in the easier parts.

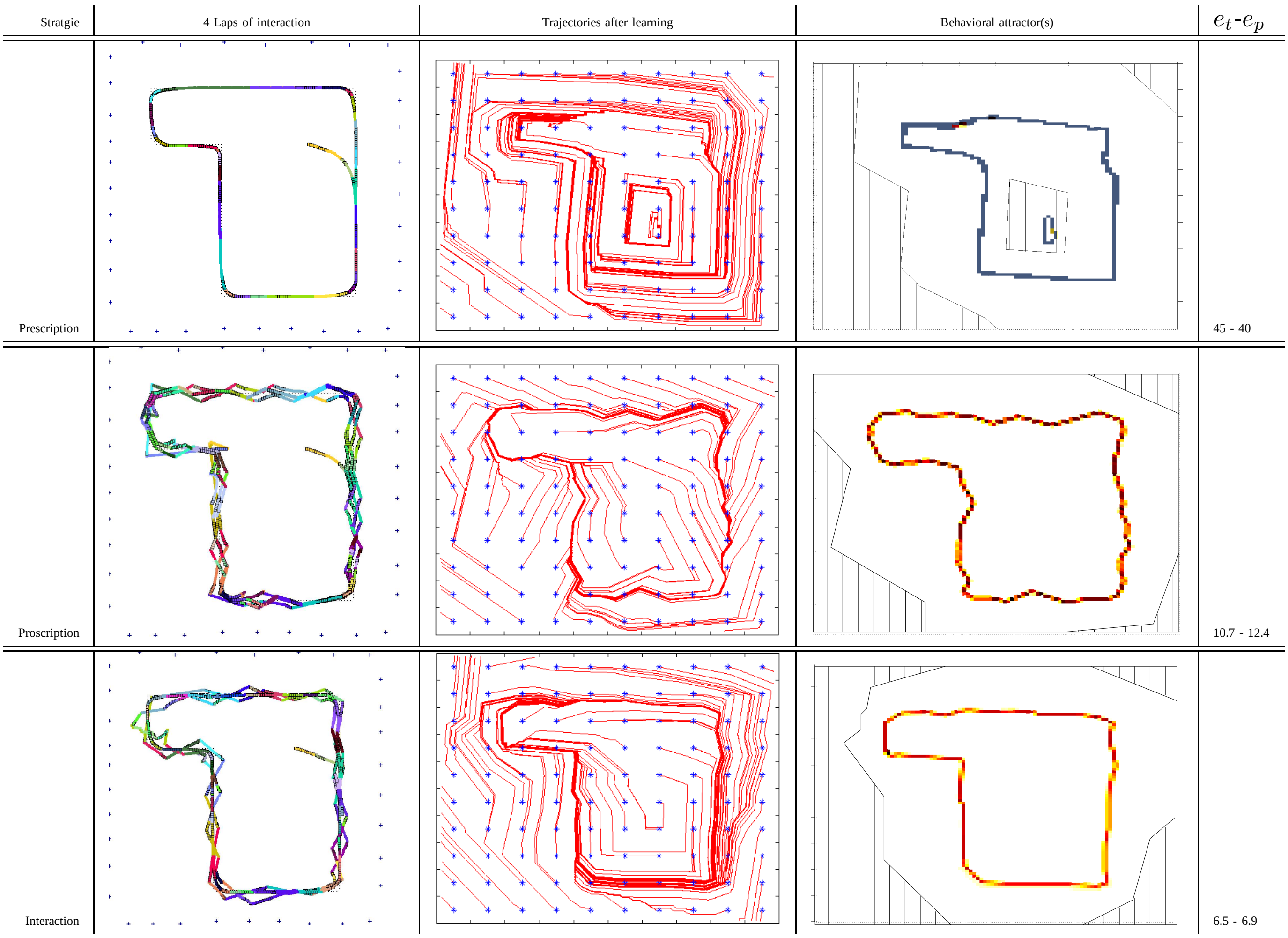


Fig. 13. Left figures show the trajectories during the four laps of training. The figures in the centre show some generated trajectories. The behavioral attractors and their attraction basins are displayed on the right figures (the attractors correspond to the mean position of the robot after a long time for different starting points and the attraction basin is deduced from the fig. of the trajectories). Each line corresponds to a given teaching strategy. In the first experiment, the prescriptive teaching is simulated. The trajectories either diverge or converge towards a bad attractor. For this attractor, $e_t = 45$ and $e_p = 40$. A second parasitic attractor has also been created. In the second experiment, the proscriptive teaching is simulated. The simulated teacher never shows the precise trajectory to the robot. The program 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 to 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.

B. Effects of the interaction strategy

The heart of this section aims at demonstrating the interest of a real human interaction of guidance as opposed to a predefined strategy such as a purely proscriptive or prescriptive training.

1) Expected results:

The proposed PerAc architecture for local navigation enables a teacher to specify a task to a robot. Even if the communication is based on a very simple media, different strategies may be adopted by the teacher to interact with the robot. The teacher may perfectly guide the robot corresponding to a *prescriptive teaching* or on the contrary, adopt a *proscriptive teaching* consisting in correcting the robot when it is too far from the centre of the trajectory. This opposition between a prescriptive strategy and a proscriptive strategy reminds the opposition between an objectivist and a constructivist approach of the autonomy, pointed out in [76]. In both cases, the robot should be able to extract the information and to use it as well as possible. The result of the experiment illustrated by fig. 13 highlights that both kinds of learning are necessary to obtain a more accurate behavioral attractor. The teacher must let its robot commit errors to obtain a convergent behavior and he must also show the precise trajectory to refine the centre of the attraction basin. If the teacher only adopts one of two strategies, the resulting behavior is expected to be worst than if both strategies are used. An interesting point is that the course of the interaction with the robot should logically imply both kinds of learning. Based on the same experimental conditions described in the previous section, these expected results are validated.

2) Experimantal validation:

Let us first consider a prescriptive teaching (fist line of fig. 13). As the teacher always does the same action in the same places without observing the robot behavior, he never knows if the robot learns or if it is able to reproduce the behavior. Hence, neither the teacher nor the robot knows if the resulting behavior is correct. Since no interaction has really occurred, and no error has been committed, the algorithm is not able to efficiently generalize: the created dynamics has two attractors: some starting point can lead either to a convergent behavior (but the generated trajectory is quite unsatisfying), or to a parasitic fixed point (in the middle of the environment). Although, this strategy enables the robot to learn the best movement in the centre of the trajectory, the resulting attractor is bad because the robot does not know what to do when it escapes from the trajectory.

The other strategy the teacher can adopt is to correct the robot when it is too far from the centre of the trajectory, corresponding to a proscriptive teaching (second line of fig. 13). Hence the robot oscillates from a border of the allowed road to the other. This strategy has the advantage for the teacher to directly evaluate the precision of the learning by observing the errors of the robot. Moreover the locations of the place-action associations surround the precise trajectory, leading to a real convergence towards the centre of the trajectory. The drawback is that the robot does not stabilize on the precise trajectory but oscillates around it. Fig. 13 illustrates the oscillating effects of the sole proscriptive teaching. This figure also shows that the approximation of the dynamics no longer has any erroneous parasitic attractor and that the generated attractor is far more accurate (see the measures e_t and e_p, divides by 4 as compared to the result of the presctitive teaching).

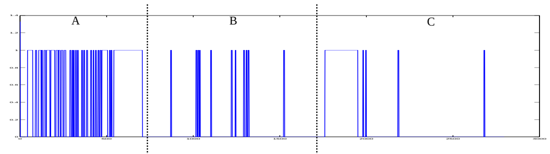


Fig. 14. Rythm of the interaction between a human teacher and the robot during an experiment like the last one of the fig. 13 (but the number of training laps was not constrained). The graph shows when the human presses the button. The phases of proscriptive teaching correspond to the Dirac pulses. The phases of prescriptive teaching correspond to the longer step. Three periods emerge: during the period A, corresponding to the beginning of the interaction, the correction frequency is high: the teacher has to be directive since the robot knows nothing. The period B is characterized by an alternation of correction phases and observation phases. The period C corresponds to the final step of the learning: the teacher tries to finalize the training first by a long prescriptive phase, and then by selecting particular proscriptive orders.

The simulations of the prescriptive as well as the proscriptive strategies are in fact adhoc processes of guidance which does not require any human intervention. If a human is asked to decide when to correct the robot by pressing a button (the simulated human guidance is activated as long as the button is pressed, and the robot realizes the learned behavior otherwise), both kinds of learning will naturally emerge from the interaction (see the last line of fig. 13). During the natural course of the interaction, the teacher oscillates between precise demonstrations of the trajectory (prescriptive teaching), observation of the robot behavior and proscriptive corrections as shown in fig. 14 which illustrated the rhythm of the interaction (the number of laps of training in this experiment has not been constrained). The human and the robot really interact by means of a non-verbal, non-symbolic language based on the actions (imposed by the teacher and reproduced by the robot). The fig.14 illustrated the rhythm of the interaction. The teacher alternates between prescriptive and proscriptive phases. As a result, we can see that the generated trajectories are more precise than when the single proscriptive teaching was used. Both strategies have actually complementary properties and occur successively during the real interaction. The proscriptive teaching enables to create the border of the attraction basin guaranteeing a convergence towards the centre of the trajectory, whereas the prescriptive teaching enables to precisely dig the centre of the attraction basin.

C. Experiments with real robots

We present here results in real indoor and outdoor environments in order to highlight the usability of the system and to show the expected course of the real time interaction.

The experiments proposed here show the accuracy of our approach in real environment. The indoor experiments (see

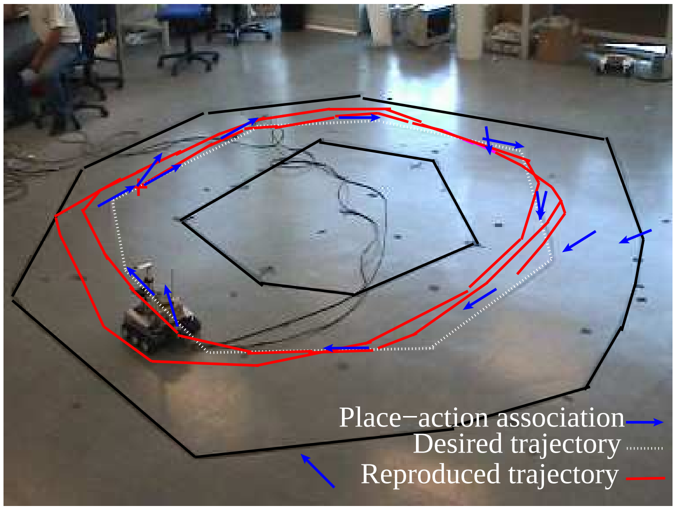


Fig. 15. Indoor experiment: the robot is guided by a human operator. Three laps are sufficient to train the robot to perform the task within the road defined by the black border (not visible from the robot).

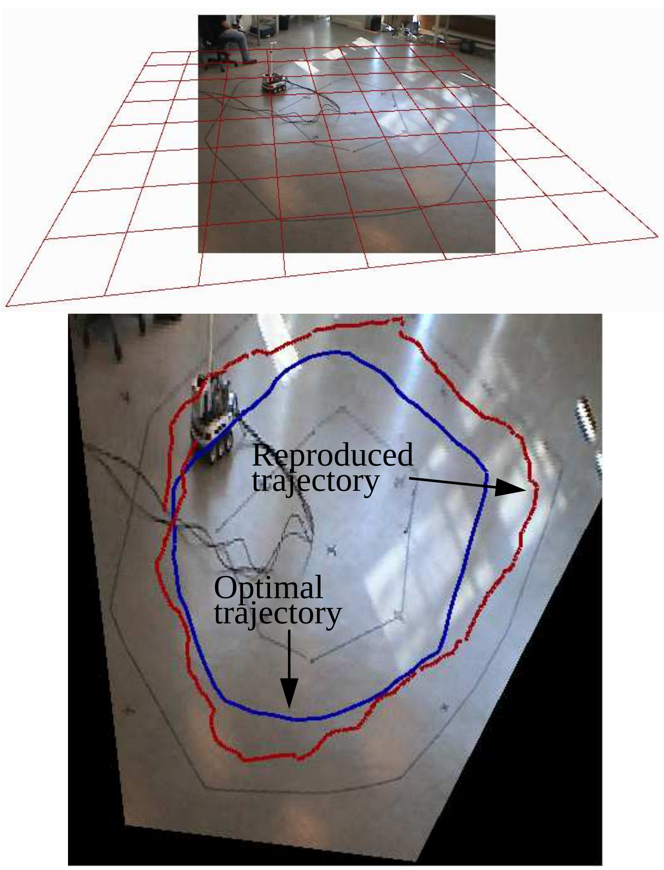


Fig. 16. Measure of an indoor trajectory. The perspective effect of the camera used to record the experiments is first corrected. Then, the user specifies the optimal trajectory. The tracking of the robot in the corrected image enables to compute e_t and e_p. In this experiment, $ep = 23cm$ and $et = 26cm$ (1 square of the grid represents 0.75 m)

fig. 15 and 16) demonstrate in favorable conditions (constant artificial light, horizontal ground, various and numerous visual landmarks) that it is quite easy to train a robot to perform a sensory-motor task. In the experiment of fig. 15, three laps were sufficient for the robot to learn a convergent behavior. The precision could have been further enhanced by guiding the robot on more laps. In a second experiment, we tried to measure e_t and e_p experimentally. In order to extract the real trajectory of the robot and compare it with the desired trajectory, a visual tracking system is used. Fig 16 illustrates the tracking and the perspective correction used to measure e_t and e_p. In this experiment, $e_p = 23cm$ and $e_t = 26cm$.

Outdoor experiments are far more difficult to analyze due to the constraints of the natural rought environments. The size and the nature of the experimental environment avoided us to record the precise trajectory: two or three synchronized camera would have been necessary, and the GPS does not work in such a ”urban canyons”. Moreover, a stabilized platform using two accelerometers was necessary. The robot camera and its magnetic compass were mounted on this stabilized platform (see fig. 4.b) to deal with the non-planarity of the ground leading, otherwise, to errors in compass and vision measurements. This platform enables to limit the effects of a non-planar ground on the sensory measurements. For outdoor experiments, we had to improve the robustness of our vision system to deal with high and quick variations of the luminance conditions, when the robot camera captures buildings directly illuminated by the sun as compared to shadowed area. We had to develop an exposure-time and gain adaptor to control the parameter of our firewire CCD camera. Moreover, the sonar system of the pioneer AT was almost unusable since it was unable to differentiate a natural slope of the road from the walls and since it detected the long grass as an obstacle. In spite of these difficulties, we succeeded in teaching an accurate trajectory to the robot according to the expected theoretical precision, with only two laps of proscriptive teaching (see fig. 17). Only 14 places were learned which is extremely low as compared to the environment size.

V. Discusion

The choice of our adaptive one shot learning (section III) is questionable since it does not aim at guaranteeing an optimal policy; yet it offers a lot of advantages. The one shot learning creates a first coarse approximation of the desired behavioral dynamics which can be directly used. Hence, the teacher can directly see the consequences of its guidance when a place-action association is being learned. The one shot learning has also the property to be instantaneously usable by the robot, giving a real feed-back to the human on the effect of its actions on the learning. As a simple one-shot association does not enable to refine the behavioral dynamic, it is necessary to modify the sensory-motor learning rule in order to take into account a possible adaptation. The adaptation capability is also crucial in order to deal with an imprecise guiding. During the crossing of a place, the robot integrates the performed movements, without wondering if the movements are performed actively or passively (*i.e.* if the movements are decided by the robot or imposed by the teacher). When the robot enters another place, the integrated movement can be used to adapt the learned movement associated to the previous place. Hence

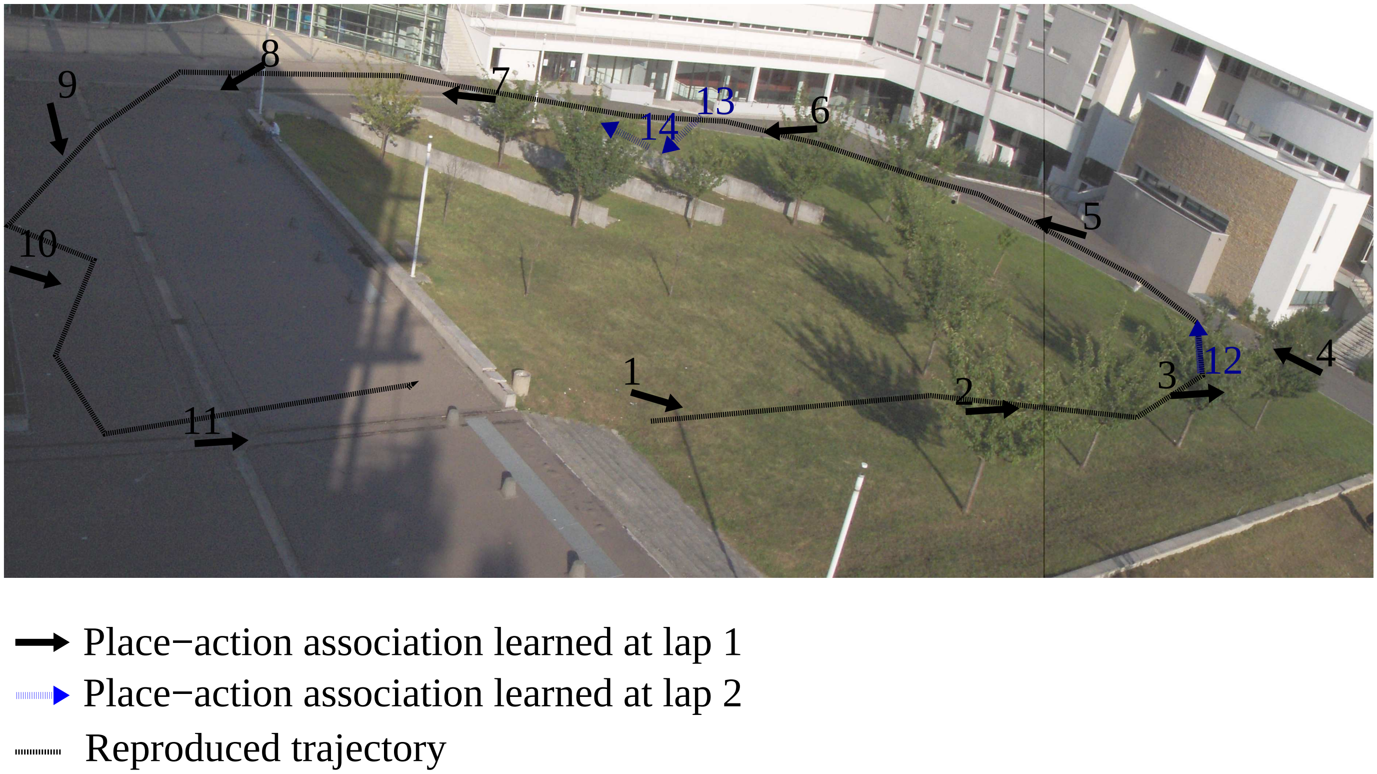


Fig. 17. Outdoor experiment of interactive teaching of a visual path. The 200 m are covered in 9 mn. The architecture is split on four processors but the speed limitation is due to the low level drivers of the robot actuators. Otherwise the robot speed could be higher.

the corrections provided by the teacher enable to refine the dynamic whereas the autonomous movements of the robot reinforce the learned dynamics. Such an adaptive learning process enables to approximate the behavior as precisely as the spatial partitioning and the behavior of the teacher allow it.

A crucial problem of the interaction between humans and robots is that the human can never know if the task was correctly 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? Since the teacher corrects the robot, he can not know what the behavior would have been if no correction had been given. Hence, a single prescriptive teaching is insufficient to produce a constructive interaction as previously illustrated. The teacher has to evaluate the robot behavior by both prescriptive and proscriptive teaching. Moreover, during such a HRI, a real communication based on actions emerges [74]: the teacher communicates by controlling the joystick, and the robot communicates by behaving according to its learned sensory-motor dynamics. The interaction is composed of three different phases which alternate: prescriptive teaching phases, proscriptive teaching phases and observation/demonstration phases. The three phases alternate in time and space according to the evolution of the teaching, defining an interaction rhythm. We pointed out (section IV) that the richness of a real HRI (human-robot interaction) acts as a cognitive catalyst, enhancing the precision of the reproduced behavior as well as its functioning domain (size of the attraction basin) [5], as opposed to a pre-determined guidance strategy such as a prescriptive guidance commonly referred as programming by demonstration (which is in fact far from being an interactive learning process since the behavior of the robot does not modify the teacher behavior) or a proscriptive strategy consisting of correcting the robot when it commits errors (which is the first step towards an interactive learning, since the behavior of the robot influences the teacher behavior, modifying the robot behavior, and so on and so on...), The system was finally validated by experiments in real indoor and outdoor environments. The experiments prove that the system is really mature and could be used by naive user in the human world for patrolling in an a priory unknown environment.

Two major problems remain: the control of the end of the interaction and the possibility that permanent environmental changes occur. Indeed, the teacher can finally be satisfied with the robot behavior or not. If the trainer is not satisfied with the robot behavior whereas the learning has already converged, this must mean that the desired precision is not reachable by 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 could communicate its desire to progress, or even guide the teacher in these states. The interaction could be more constructive if the robot had the possibility to disobey the teacher in known states or expresses its need of help in unmastered states. Such variations of the behavior will constitute another excellent feed-back for the teacher 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 it to progress, or if its predictions are normal according to the current situation. We currently work on a progress-based approach derived from [13], [77] for the meta-control of the learning, aiming at giving self-evaluation capabilities to the robots. In [78], [9], [79], we proposed a progress-based neural architecture which is proved to provide the robot with the capability to detect phases of progress, phases of stagnation, and novelty. Novelty detection leads the robot to re-adapt its erroneous learning.

We also want to investigate how to give to the robot a specific behavior according to its self-evaluation of its mastery of the task in order to enrich the interaction and to speed up the knowledge transfer. In unmastered situations, the robot could use repair strategies to get back the attention of the teacher, by means of a particular behavior (oscillation, stop, looking toward the teacher ...) [80], or by means of a more understandable media as an expressive robot head [81], [82]. Seeing these behavioral oscillations, the teacher should interact with the robot by giving him the correct orientation, providing additional examples for the learning. In mastered states, the robot could become curious by choosing to not realize the learned behavior and to disobey its teacher in order to find less mastered states in which it can still progress (a communication based on the expression of emotional states could once again be very pertinent). Indeed, this could lead the robot towards states it would not have experimented if it had performed what it had learned, or if it had perform the predictable actions imposed by its teacher. Obviously, auto-evaluation capabilities also appear as an excellent starting point to deal with permanent environmental changes or morphological changes of the robot: self-evaluation capabilities could more easily lead to consider re-learning strategy in case of such permanent changes.

VI. Conclusions and perspectives

In this paper, we addressed the problem of the interactive teaching of a sensory-motor navigation task to a mobile robot. The proposed sensory-motor learning rule enables the robot to associate newly learned places with the current action by means of a classical WH learning rule and to refine the learned behavior by merging the learned movement in each place with the performed movement by means of a delayed WH learning rule. By triggering the recruitment of new places according to the sensory-motor error, the proposed generalization of the PerAc architecture adapts the partitioning of the environment to the complexity of the task to learn. The use of a joystick to teach the robot, in spite of its simplicity, creates a real human-robot interaction with the emergence of a *dialogue* based on actions. We proposed accuracy measures and highlighted the fact that human-robot interactions can catalyze the learning and speed up its convergence. Experiments in both indoor and outdoor environments were presented in order to evaluate the performance of the whole system for the control of a real robot.

Future works will focus on the comparison of sensory-motor strategies versus planning strategies for the interactive learning of an arbitrary path and the control of its reproduction. In our complete biological model of the navigation, neurons in the hippocampus proper (CA1/CA3 regions) learn and predict transitions between successive multi-modal states [41]. A cognitive map performs a latent learning of the spatial topology of the environment [67] and can be used to compute a plan of actions to reach an arbitrary goal [64]. The system has been recently validated in an experiment of a long random exploration (45mn, 3000 steps of the place-cell architecture), in a real indoor environment [38]. The experiment highlights the capability of the system to predict transitions of places, to latently build the cognitive map of the learned transitions, and thus to plan trajectory to particular goals specified by a simple reinforcement in the location of the goal at the end of the exploration. The influence of our progress-based meta-controller [78], [79] will be evaluated at every level of this architecture. We will also study how an agent can autonomously detect it is not really doing what it aims at doing. We will wonder how an emotional system could be used as a second order controller [81] to adjust the shape of the attraction basins provided by the sensory-motor or the planning systems when the behavior becomes incorrect.

Finally, we are currently addressing the problem of the building of a single architecture, allowing the robot to deal with spatial as well as temporal modalities (place-action and duration-action stategy), in navigation as well as in robotics arm manipulation [39]. Our perspective is to build a merged control architecture for applications in which navigation and object manipulation are considered. A simple example could be to imagine a robot that must be able to use the door handles or to press elevator buttons to achieve its mission. However, this kind of missions also imply the incorporation of object recognition, visual affordance detection [1], [83], [84], [85], [86] and more sophisticated mechanisms to understand the natural and/or human world. The adaptation of our system on UAVs (Unmanned Aerial Vehicules) is also currently studied.

Movies of the experiments presented in fig. 6, 15 and 17 are available on the website of the authors and on:
http://www.etis.ensea.fr/˜neurocyber/giovannangeli/Home.html

Ackowledgment

This work is supported by the DGA (Délegation Générale pour l'Armement: procurement contract 04 51 022 00 470 27 75), the Institut Universitaire de France, the Feelix Growning European project (FP6 IST-045169). The Visiontrain project (FP6 MRTN-CT-2004-005439) supported part of the results dissemination. We particularly thanks G. Désilles for its investment in our research, J.P. Banquet for the biological counterpart of the model, and all the members of the Neurocybernetic Team for their daily support and the tremendous discussions on cognitive robotics and more.

References

[1] J. Gibson, *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.
[2] F. Varela, *Invitation aux sciences cognitives*. Seuil, 1996.

[3] J. von Uexküll, *Mondes animaux et monde humain*. Gonthier, Paris, 1966.

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

[5] C. Giovannangeli and P. Gaussier, "Human-robot interactions as a cognitive catalyst for the learning of behavioral attractors," in *16th IEEE International Symposium on Robot and Human Interactive Communication 2007*, Jeju, South Korea, 2007, pp. 1028–1033.

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

[7] F. Michaud, J.-F. Laplante, H. Larouche, A. Duquette, S. Caron, D. Letourneau, and P. Masson, "Autonomous spherical mobile robot for child-development studies," *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, vol. 35(4), pp. 471 – 480, 2005.

[8] S. Thrun, "Robotic mapping: A survey," in *Exploring Artificial Intelligence in the New Millenium*, G. Lakemeyer and e. Nebel, B., Eds. Morgan Kaufmann, 2002.

[9] C. Giovannangeli, "Navigation autonome en environnement intrieur et extrieur: apprentissage sensori-moteur et planification dans un cadre interactif," Ph.D. dissertation, Universit de Cergy-Pontoise, 2007.

[10] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, "Real-time visual loop-closure detection," in *2008 IEEE International Conference on Robotics and Automation (ICRA2008)*, 2008.

[11] J. Schmidhuber, "Curious model-building control systems," in *Proc. Int. Conf. on Neural Networks*, vol. 2. IEEE Press, 1991, pp. 1458–1463.

[12] J. Schmidhuber, J. Zhao, and M. Wiering, *Learning to learn*. Kluwer, 1997, ch. Reinforcement learning with self-modifying policies, pp. 293–309.

[13] 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.

[14] P.-Y. Oudeyer, F. Kaplan, V. Hafner, and A. Whyte, "The playground experiment: task-independent development of a curios robot." in *Proc. of the AAAI Spring Symposium Workshop on Developmental Robotics*, D. Bank and e. Meeden, L., Eds., 2005, pp. 42–47.

[15] C. Tovey and S. Koenig, "Improved analysis of greedy mapping," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2003, pp. 3251–3257.

[16] H. Andreasson, A. Treptow, and T. Duckett, "Localization for mobile robots using panoramic vision, local features and particle filter," in *IEEE Int. Conf. on Robotics and Automation (ICRA 2005)*, Barcelona, Spain, 2005.

[17] 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.

[18] C. Giovannangeli, P. Gaussier, and G. Désilles, "Robust mapless outdoor vision-based navigation," in *Proc. of the 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2006)*, Beijing, China, 2006, pp. 3293–3300.

[19] P. Andry, P. Gaussier, S. Moga, J. Banquet, and J. Nadel, "Learning and communication in imitation: An autonomous robot perspective," *IEEE transactions on Systems, Man and Cybernetics, Part A*, vol. 31, no. 5, pp. 431–444, 2001.

[20] P. Gaussier, S. Boucenna, and J. Nadel, "Emotional interactions as a way to structure learning," in *Epigenetic Robotics*. Lucs, 2007, pp. 193–194. [Online]. Available: http://publi-etis.ensea.fr/2007/GBN07

[21] 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.

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

[23] S. Thrun and T. Mitchell, "Lifelong robot learning," *Robotics and Autonomous Systems*, vol. 15, pp. 25–46, 1995.

[24] G. Hayes and J. Demiris, "A robot controller using learning by imitation," in *Proc. 2nd Int. Symposium on Intelligent Robotic Systems*, Grenoble, France, 1994, pp. 198–204.

[25] K. Dautenhahn, "Getting to know each other : artificial social intelligence for autonomous robots," *Robotics and Autonomous Systems*, vol. 16, pp. 333–356, 1995.

[26] C. G. Atkeson and S. Schaal, "Robot learning from demonstration," in *Machine Learning: Proceedings of the fourteenth International Conference (ICML '97)*, 1997, pp. 12–20.

[27] P. Gaussier, S. Moga, M. Quoy, and J. Banquet, "From perception-action loops to imitation processes: a bottom-up approach of learning by imitation," *Applied Artificial Intelligence*, vol. 12, no. 7-8, pp. 701–727, Oct-Dec 1998.

[28] A. Billard and M. J. Mataric, "Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture," *Robotics and Autonomous Systems*, vol. 37, no. 2-3, pp. 145–160, 2001.

[29] A. Alissandrakis, C. Nehaniv, and K. Dautenhahn, "Imitation with alice: learning to imitate corresponding actions across dissimilar embodiments," *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, vol. 32(4), pp. 482 – 496, 2002.

[30] 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.

[31] M. Nicolescu and M. J. Mataric, "Task learning through imitation and human-robot interaction," in *Models and Mechanisms of Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*. Kerstin Dautenhahn and Chrystopher Nehaniv Eds, 2005.

[32] A. Billard, Y. Epars, S. Calinon, S. Schaal, and G. Cheng, "Discovering optimal imitation strategies," *Robotics and Autonomous Systems, Special Issue: Robot Learning from Demonstration*, vol. 47, no. 2-3, pp. 69–77, 2004.

[33] S. Calinon, F. Guenter, and A. Billard, "On learning, representing and generalizing a task in a humanoid robot," *IEEE Transactions on Systems, Man and Cybernetics, Part B. Special issue on robot learning by observation, demonstration and imitation*, vol. 37(2), pp. 286–298, 2007.

[34] S. Calinon and A. Billard, "Active teaching in robot programming by demonstration," in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2007)*, 2007.

[35] ——, "What is the teacher's role in robot programming by demonstration? toward benchmarks for improved learning," *Interaction Studies. Special Issue on Psychological Benchmarks in Human-Robot Interaction*, vol. 8:3, 2007.

[36] 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.

[37] P. Gaussier, A. Revel, C. Joulain, and S. Zrehen, "Living in a partially structured environment: How to bypass the limitation of classical reinforcement techniques," *Robotics and Autonomous Systems*, vol. 20, pp. 225–250, 1997.

[38] C. Giovannangeli and P. Gaussier, "Autonomous vision-based navigation: Goal-oriented action planning by transient states prediction, cognitive map building, and sensory-motor learning," in *Proc. of the 2008 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2008)*, Nice, France, 2008.

[39] M. Lagarde, P. Andry, P. Gaussier, and C. Giovannangeli, "Learning new behaviors : Toward a control architecture merging spatial and temporal modalities," in *Workshop on Interactive Robot Learning - International Conference on Robotics: Science and Systems (RSS 2008)*, 2008.

[40] 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.

[41] 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.

[42] 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.

[43] B. Cartwright and T. Collett, "Landmark learning in bees," *Journal Comp. Physiology*, vol. 151, pp. 521–543, 1983.

[44] D. Lambrinos, R. Moller, T. Labhart, R. Pfeifer, and R. Wehner, "A mobile robot employing insect strategies for navigation," *Robotics and Autonomous Systems*, vol. 30, pp. 39–64, 2000.

[45] V. V. Hafner, "Learning places in newly explored environment," in *SAB2000 Proc. supplement book, Publication of the International Society for Adaptive Behavior (Honolulu)*. IOS Press, 2000, pp. 111–120.

[46] A. Argyros, C. Bekris, S. Orphanoudakis, and L. Kavraki, "Robot homing by exploiting panoramic vision," *Journal of Autonomous Rbots*, vol. 19, no. 1, pp. 7–25, 2005.

[47] V. V. Hafner, "Cognitive maps in rats and robots," *Adaptive Behavior*, vol. 13, no. 2, pp. 87–96, 2005.

[48] L. Smith, A. Philippides, and H. P., "Navigation in large-scale environments using an augmented model of visual homing," in *From Animals to Animats 9: Proceedings of the 9th International Conference on Simulation of Adaptive Behaviour, SAB06*, J. G. Carbonell and e. J. Siekmann, Eds. Springer Berlin/Heidelberg, 2006, pp. 251–262.

[49] L. Smith, A. Philippides, P. Graham, B. Baddeley, and P. Husbands, "Linked local navigation for visual route guidance," *Adaptive Behavior - Animals, Animats, Software Agents, Robots, Adaptive Systems*, vol. 15, no. 3, pp. 257–271, 2007.

[50] R. Bellman, *Dynamic Programming*. Princeton, N.J.: Princeton University Press, 1957.

[51] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine Learning*, vol. 3, pp. 9–44, 1988.

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

[53] C. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, pp. 279–292, 1992.

[54] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol. 8, pp. 229–256, 1992.

[55] L. Baird, "Gradient descent for general reinforcement learning," *Neural InformationProcessing Systems*, vol. 11., 1998.

[56] S. Leprêtre, P. Gaussier, and J. Cocquerez, "From navigation to active object recognition," in *The Sixth Int. Conf. on Simulation for Adaptive Behavior SAB'2000*. Paris: MIT Press, 2000, pp. 266–275.

[57] C. Giovannangeli and P. Gaussier, "Orientation system in robots: Merging allothetic and idiothetic estimations," in *Proc of the 13th Int. Conf. on Advanced Robotics*, Jeju, South Korea, 2007, pp. 349–354.

[58] L. P. Kaelbling, M. L. Littman, and A. P. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

[59] E. Zalama, J. Gomez, M. Paul, and J. Peran, "Adaptive behavior navigation of a mobile robot," *Systems, Man and Cybernetics, Part A, IEEE Transactions o*, vol. 32(1), pp. 160 – 169, 2002.

[60] R. Munos and J. Patinel, "Reinforcement learning with dynamic covering of state-action space: partitioning q-learning," in *SAB94: Proceedings of the third international conference on Simulation of adaptive behavior : from animals to animats 3*. Cambridge, MA, USA: MIT Press, 1994, pp. 354–363.

[61] M. Arbib and I. Lieblich, "Motivational learning of spatial behavior," *Systems Neuroscience*, pp. 77–87, 1977.

[62] M. Mataric, "Integration of representation into goal-driven behavior-based robot," *IEEE Transactions on Robotics and Automation*, vol. 8(3), pp. 304–312, 1992.

[63] B. Schölkopf and H. A. Mallot, "View-based cognitive mapping and path-finding," *Adaptive Behavior*, vol. 3, pp. 311–348, 1995.

[64] A. Revel, P. Gaussier, S. Leprêtre, and J. Banquet, "Planification versus sensory-motor conditioning: what are the issues ?" in *SAB'98: From animals to animats 5*, 1998, pp. 129–138.

[65] N. A. Schmajuk and H. Voicu, "Exploration, navigation and cognitive mapping," *Adaptive Behavior*, vol. 8(3), pp. 207 – 223, 2000.

[66] H. Andreasson and T. Duckett, "Topological localization for mobile robots using omni-directional vision and local features," in *Proc. IAV 2004, the 5th IFAC Symposium on Intelligent Autonomous Vehicles*, Lisbon, Portugal, 2004.

[67] E. Tolman, "Cognitive maps in rats and men," *The Psychological Review*, vol. 55, no. 4, pp. 189–208, 1948.

[68] J. N. Tsitsiklis and B. V. Roy, "An analysis of temporal-difference learning with function approximation," *IEEE Transactions on Automatic Control*, vol. 42, pp. 674–690, 1997.

[69] S. P. Singh, T. Jaakkola, M. L. Littman, and C. Szepesvári, "Convergence results for single-step on-policy reinforcement-learning algorithms," *Machine Learning*, vol. 38, no. 3, pp. 287–308, 2000.

[70] S. J. Bradtke, "Reinforcement learning applied to linear quadratic regulation," *Advances in Neural Information Processing Systems*, vol. 11, pp. 295–302, 1993.

[71] S. Thrun and A. Schwartz, "Issues in Using Function Approximation for Reinforcement Learning," in *Proceedings of the 1993 Connectionist Models Summer School*, M. Mozer, P. Smolensky, D. Touretzky, J. Elman, and A. Weigend, Eds. Hillsdale, NJ: Lawrence Erlbaum, 1993.

[72] W. D. Smart and L. P. Kaelbling, "Practical reinforcement learning in continuous spaces," in *Proc. 17th International Conf. on Machine Learning*. Morgan Kaufmann, San Francisco, CA, 2000, pp. 903–910.

[73] J. Varela, E. Thompson, and E. Rosch, *The embodied mind: Cognitive science and human experience*. MIT Press, Cambridge, MA, USA, 1991.

[74] N. Koenig and M. J. Mataric, "Behavior-based segmentation of demonstrated task," in *IEEE International Conference on Development and Learning (ICDL)*, 2006.

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

[76] J. Stewart, "The implication for understanding high-level cognition of a grounding in elementary adaptive systems," *Robotics and Autonomous Systems*, vol. 16(2-4), p. 107116, 1995.

[77] 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.

[78] C. Giovannangeli and P. Gaussier, "Learning to navigate, progress and self-evaluation," in *6th International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, 2006.

[79] C. Giovannangeli, S. Boucenna, and P. Gaussier, "About the constuctivist role of self-evaluation for interactive learnings and self-development," in *Proc. of the 2008 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2008). Workshop: From motor to interaction learning in robots*, 2008.

[80] C. Muhl and Y. Nagai, "Does disturbance discourage people from communicating with a robot?" in *In Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'07)*, 2007.

[81] L. Canamero and P. Gaussier, *Emotion understanding: robots as tools and models*. J. Nadel and D. Muir, 2005, pp. 235–258.

[82] M. Ogino, A. Watanabe, and M. Asada, "Mapping from facial expression to internal state based on intuitive parenting," in *Proceedings of the Sixth International Workshop on Epigenetic Robotics*, 2006, pp. 182–183.

[83] M. Steedman, "Formalizing affordance," in *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*, W. D. Lawrence Erlbaum, Ed., 2002, p. 834839.

[84] A. Chemero, "An outline of a theory of affordances," *Ecological Psychology*, vol. 15(2), p. 181195, 2003.

[85] E. Sahin, M. Cakmak, M. Dogar, E. Ugur, and G. Ucoluk, "To afford or not to afford: A new formalization of affordances towards affordance-based robot control," *Adaptive Behavior*, vol. 15(4), pp. 447–472, 2007.

[86] E. Rome, L. Paletta, E. Sahin, G. Dorffner, J. Hertzberg, R. Breithaupt, G. Fritz, J. Irran, F. Kintzler, C. Lrken, S. May, and E. Uur, "The macs project: An approach to affordance-inspired robot control," *Lecture Notes in Computer Sciences*, vol. 4760, pp. 173–210, 2008.

Christophe Giovannangeli is postdoctorant at the Technion in Israel, in the Computer Science Department. He obtained its PhD at the Cergy Pontoise University in France in 2007, where he worked with the Neurocybernetic team of the Image and Signal processing Lab (ETIS). Its current research interest are: pursuit-evasion games in presence of obstacles for autonomous UGV and UAV, landmarks selection and recognition, online environnement learning, action planning, robot behavior teaching by human-robot interactions and bio-inspired robotics.

Philippe Gaussier s Professor at the Cergy Pontoise University in France and leads the neurocybernetic team of the Image and Signal processing Lab. He is also member of the Institut Universitaire de France. Currently his research interests are focused on the modelling of cognitive mechanisms and brain structures such as the hippocampus and its relations with cortical structures like parietal, temporal and pre-frontal areas, the dynamics of visual perception,.the development of interaction capabilities (imitation, emotions...). Current robotic applications include autonomous and online learning for motivated visual navigation, object manipulation... (place learning, visual homing, object discremination, action planning...).