Interactive Teaching for Vision-Based Mobile Robots: A Sensory-Motor Approach

Christophe Giovannangeli and Philippe Gaussier

***Abstract*—For the last decade, we have been developing a vision-based architecture for mobile robot navigation. Using our bio-inspired model of navigation, robots can perform sensory-motor tasks in real time in unknown indoor as well as outdoor environments. We address here the problem of autonomous incremental learning of a sensory-motor task, demonstrated by an operator guiding a robot. The proposed system allows for semisupervision of 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, used first in a real experiment and then in a simulated experiment, demonstrate how a real interaction between the teacher and the robot influences the learning process.**

***Index Terms*—Cooperative systems, intelligent robots, learning systems, mobile robots, navigation, robot vision systems.**

I. Introduction

TASK specification in autonomous robotics has attracted increasing interest in recent years. It is now acknowledged that autonomous mobile robots should be designed with minimal prior knowledge about the tasks to be performed so that the robot can adapt to unpredictable situations that characterize the dynamic nature of real environments. The robots should also develop their skills via interactions with their physical and social environment, where they experience sensory-motor interactions [1], allowing for the emergence of their own cognition and the building of a subjective enacted world [2], also called *Umwelt* [3]. In this context, human–robot interactions (HRIs) are thought to be a very efficient means for specifying various tasks to a robot [4] and catalyzing its sensory-motor learning [5]. HRIs are, moreover, crucial for designing operational or social and interactive robots [6], [7]. These statements can be extended to robot–robot interactions [8]. This paper investigates the use of HRI for the learning of navigation tasks.

Manuscript received January 24, 2008; revised July 1, 2008. First published December 4, 2009; current version published December 16, 2009. This work was supported in part by Délegation Générale pour l'Armement under Procurement Contract 04 51 022 00 470 27 75, by the Institut Universitaire de France, by Feelix Growing European Project FP6 IST-045169, and by the Visiontrain Project FP6 MRTN-CT-2004-005439. This paper was recommended by Associate Editor G. C. Calafiore.

The authors are with the Neurocybernetic Team, Image and Signal Processing (ETIS) Laboratory, Cergy-Pontoise University, 95302 Cergy-Pontoise, France (e-mail: christophe.giovannangeli@gmail.com; gaussier@ensea.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2009.2033029

In [9], several problems linked to the autonomous localization and mapping of an environment are pointed out. We summarize here the main points in our approach: 1) The nature of the "noise" on the physical measurements is generally context dependent. For example, a noncalibrated panoramic camera can induce a biased error of the landmark position measurement, depending on the robot's orientation [10]. Whereas several statistical methods can deal with centered noises, it is much more difficult to detect a conditionally biased noise because such noise is characterized by a mean value and a standard deviation, which depend on the unexpected (hidden) dimensions of the robot's 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 a real-time robotics architecture for pseudoinfinite environments. 3) The algorithms must be able to handle the correspondence problem (or data association problem) to reliably determine whether two sensorial measurements taken at different time steps correspond to the same physical point in the environment [11]. For a long time, this problem has been treated as stochastic, leading the community to develop algorithms that try to reveal the hidden Markovian model of the environment. Yet, psychology, long ago, identified the ambiguous nature of perception (Gestalt theory): The so-called multistability of perception implies that a perfectly well-defined sensory stimulus can have two opposing interpretations (Necker's cubes, Rubin's figure, and other artistic creations are good examples). This ambiguous nature of perception should cause roboticists to question whether such a hidden Markovian model of the environment really does exist or it is meaningless to try to remove sensorial ambiguities. 4) The dynamic nature of the environment induces environmental changes. Localization cues used by a robot may disappear during its lifetime. When the environment changes, 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 relearning strategies, enabling them to adapt their knowledge to the environmental changes before their behaviors become completely irrelevant. Finally, 5) the robot is confronted with 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 metalearning theory, for instance, claimed early on that a smart selection of the prototypes for learning can increase the developmental speed of the robot [12], [13]. Recent works insist that selecting the action that maximizes the learning progress makes the robot curious and enables it to develop faster [14], [15]. The complexity of greedy

mapping algorithms in deterministic environments was studied in [16]. Moreover, developers are confronted with a tradeoff between learning speed and system reliability: The accuracy of fast incremental algorithms strongly depends on the quality of the sensory measurements, whereas statistical methods, which are asymptotically more accurate, require numerous examples of the environment to build a consistent map. Despite these limitations, simultaneous localization and mapping (SLAM) methods exhibit impressive performance in long-term navigation when coupled with visual recognition systems to help them deal with the correspondence problem and the environmental changes [17]. Nowadays, it is possible to map large indoor environments with monocular, stereoscopic, or catadioptric vision systems, although the scaling factor for larger and less controlled environments raises some rarely addressed questions: size explosion of the internal representations, fusion of isolated maps, unreliability of the wheel-based odometry on a rough terrain, and the ground planarity hypothesis. The sensory-motor system for mobile robot navigation that is the focus of this paper has already proved its efficiency with regard to some of these drawbacks [18], [19].

In the following, we will propose a visual navigation architecture that is bootstrapped for task specification by imitation and 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 such as visual path following or homing, and the robot has to reproduce it. The robot will be guided by a joystick that will be used as an approximation of an imitation process (other works in our laboratory focus on this aspect [20], [21]). In [22], 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). Nicolescu and Mataric also point out that *acting* can provide a basis for a nonverbal 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 its teacher questions has already been evaluated, for example, in the collaborative control in [23]. *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 a symbolic language that the robot understands. As a general rule, task specification is performed at a very high symbolic level (as highlighted in [24]), under the autocratic rule of the teacher. Other recent works focus on this kind of cooperation [25] in navigation. However, most of the learning systems based on imitation need to separate the learning phases and the performance phases. Yet, constraints on lifelong learning [26] imply that the robot must be able to learn while freely evolving in the world. A less unilateral process for task specification could emerge from an interaction 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 to be of interest in machine learning and, more specifically, in robot skill learning, as illustrated by various studies in the last 15 years [27]–[34]. Theoretical studies have also been undertaken, as in [35], which presents a general formalism for performance metrics on humanoid imitation tasks and illustrates the need for a general framework in order to evaluate the relative accuracy of different algorithms. However, the imitation as a real dynamic and continuous HRI has hardly ever been stressed (rare examples are [20], [21], and [33]). Most of the imitation learning and teaching methods are composed of a demonstration phase (the learning) and a performance phase (the reproduction of the knowledge). Rarely, however, has the imitation been treated as a real dialogue based on a language of actions between the robot and the teacher, alternating between learning and performance phases. In [36]–[38], 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's arms in order to demonstrate the task. The authors point out that the recording of the teacher's demonstration does not take into account the embodiment of the robot, whereas the passive execution of the task by the robot during the learning phases does and, hence, is far more pertinent. Although the authors insist that the observation of the performed task plays a role in helping the teacher understand the robot's skills and prepare the subsequent demonstration, the proposed system does not benefit from the intervention of the teacher during task realization.

Indeed, in the context of interactive teaching, learning and demonstration phases ought to be combined in order to provide rich and natural communication that could improve the development of the robot's skills: By imitating a teacher, the robot could test the behavior that has to be learned. By *acting and reacting* to the teacher's orders, the robot should freely exhibit its mastery of the task while improving its learning [5]. At the same time, observing the robot's 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 to be nonverbal nonsymbolic communication, we claim that it is, nevertheless, a very rich type of communication [20] that is able to catalyze the robot's learning. In such an interactive context, a strong autonomy of decision, as well as a strong autonomy of learning, is necessary. As humans are involved, rapidity, precision, and adaptation of the learning are also required.

This paper first presents our robots and the visual systems for the creation of a continuous state space. Then, we propose a bootstrap[1] for the perception–action (PerAc) architecture [39] that enables the semisupervised learning of a sensory-motor behavior (a visual path and a homing behavior). The pairs of architecture equations allow for the environmental partitioning to be adapted to the complexity of the task. The system does not separate learning and performance 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 two complementary accuracy measures that are used to compare the performed trajectory to the optimal one. The importance of the interaction between the robot and the teacher during learning, particularly with regard to adopting a proscriptive teaching

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

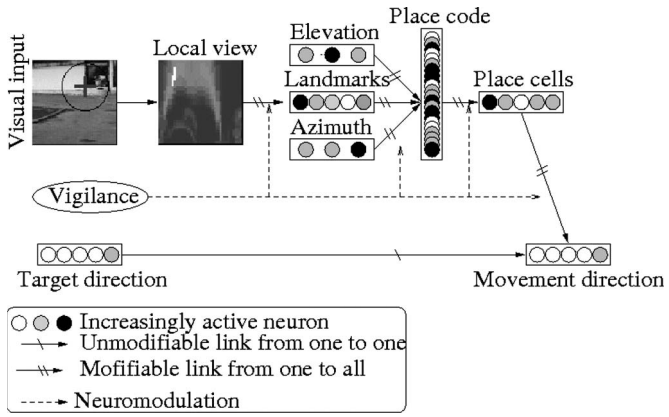

Fig. 1. Block diagram of the architecture. Our architecture for place recognition is composed of a visual system that focuses on points of interest and extracts small images in log-polar coordinates (called local views), recognized as landmarks (see Fig. 2). Next, a merging layer compresses the *what* 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 executed when the robot recognizes the place. The vigilance signal triggers waves of one-shot learning of the landmarks related to the current location, of the current place code, and of the current place and the current place–action association.

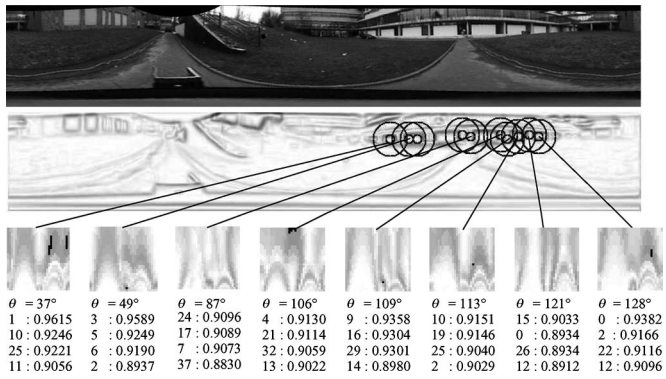

Fig. 2. Illustration of the landmark extraction mechanism: The gradient of a panoramic image is convolved with a difference-of-Gaussian filter. The local maxima of the filtered image correspond to (center of the circles) points of interest. Here, the first eight focal points are displayed. The system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

strategy allowing the robot to commit its own errors, will be experimentally illustrated using the proposed measures.

II. Methods and Materials

Among the various methods for creating spatial behaviors, the PerAc architecture [39] has been demonstrated to be particularly well adapted to online sensory-motor learning. A PerAc architecture may underlie many various skills in mobile robotics: guidance [40], local navigation in indoor [18] and outdoor environments [19], planning [41], and reproduction of a temporal sequence of actions [30], as well as in the control of actuators with multiple degrees of freedom (robotic arm control [33], [42] and 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 [43], in order to perform local navigation tasks [18], [19].

Fig. 1 shows the visual processing chain for 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 merging *what* information and *where* information provided by the visual system, which extracts local views in log-polar coordinates, centered on points of interest. Fig. 2 shows the autonomous landmark extraction mechanism.

The built-in generalization capability of the system has a remarkable property (see [43] for more details). To summarize, a place cell encoded in location A responds maximally in A and creates a decreasing place field around A over a wide area. In the experiment in Fig. 3, the robot learns 5×5 positions regularly located in a classic workroom [Fig. 3(a)]. Fig. 3(b) shows the created place field for each place cell in the whole environment, corresponding to a localization gradient. A simple approximation of a place cell activity can correspond to a noisy Gaussian curve

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

with $p_{x_l}(t)$ representing the activity in $x(t)$ of the place cell encoded in x_l, σ expressing the extent of the place field which is linked to the distance of the landmarks, and $\epsilon^P(t)$ representing the noise induced by the uncertainty of the azimuth measurements, the camera discretization, and the dynamic 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 each time all the place cell activities are less than a given threshold. As a mathematical consequence of *what* and *where*, the shape of the place fields is homothetic with the shape of the environment [42], [43] (i.e., the place fields extend with the distance to the landmarks). With regard to the problem of the size of the internal representations, our system is particularly interesting in that it builds its own metrics based on the azimuthal shifts of the landmarks and their recognition levels. 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 place cell) [43]. The computational load and the memory requirements have been proved to be a linear function of the number of learned landmarks [10]. Hence, the learning of a loop in a large outdoor environment requires as much computation load and memory as the learning of a loop in a smaller indoor environment (see experiments in Section IV). To the best of our knowledge, very few algorithms exhibit such a property. Moreover, neither Cartesian nor topological map building is needed for the localization since the world acts as an outside memory [45]. As long as the learned features of a location persist in its neighborhood, the robot is able to self-localize without map building.

The addressed problem in this paper concerns a more general class of algorithms that are based on place recognition and can lead to adaptive environmental paving. For example, GPS measurements, a triangulation system via external

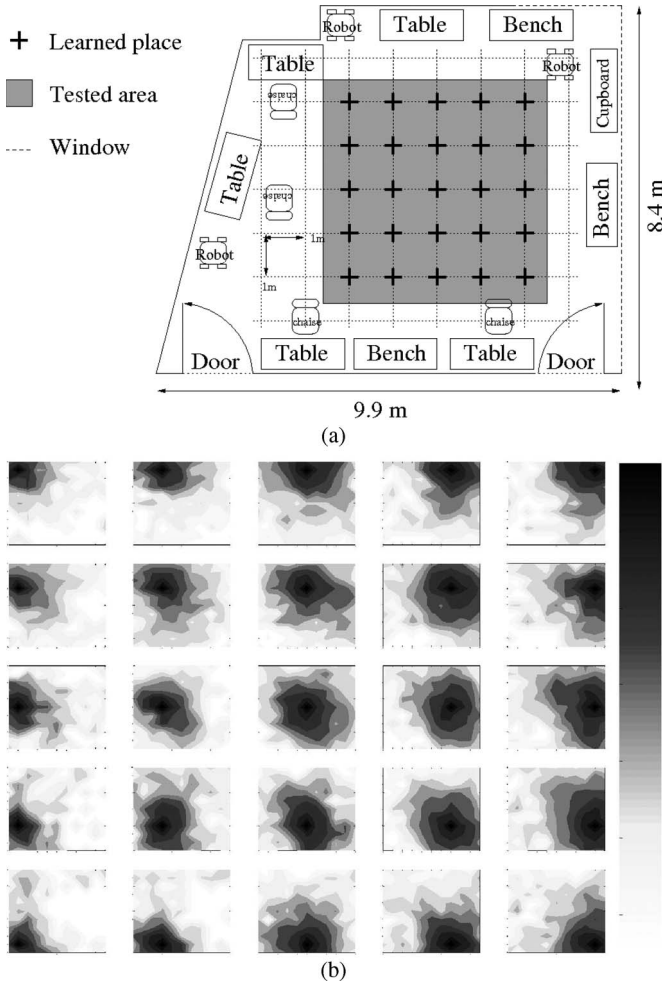

Fig. 3. (a) Workroom used in the experiment in (b). Twenty-five places are regularly learned and tested. (b) Activity of 5×5 place cells regularly encoded in the workroom in (a). The competition between all the place cells leads to the paving of the environment.

references, classical SLAM, vision-based SLAM, or topological approaches provide the information that is necessary for using 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 place cell). Unfortunately, even if the robot could maintain a given direction as long as the recognition level increases, a serious 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 insect navigation [46], [47], could be used to avoid pure 1-D hill-climbing methods. As actions are directly computed rather than learned (although learning them is possible [48]), 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 places and of place reaching (also called milestone points in [49]) or on the problem of the cognitive mapping of the learned locations [50]. Finally, the question of the robustness evaluation has rarely been raised [51]. Nevertheless, recent studies [52], [53] propose an improved version of the average landmark vector algorithm

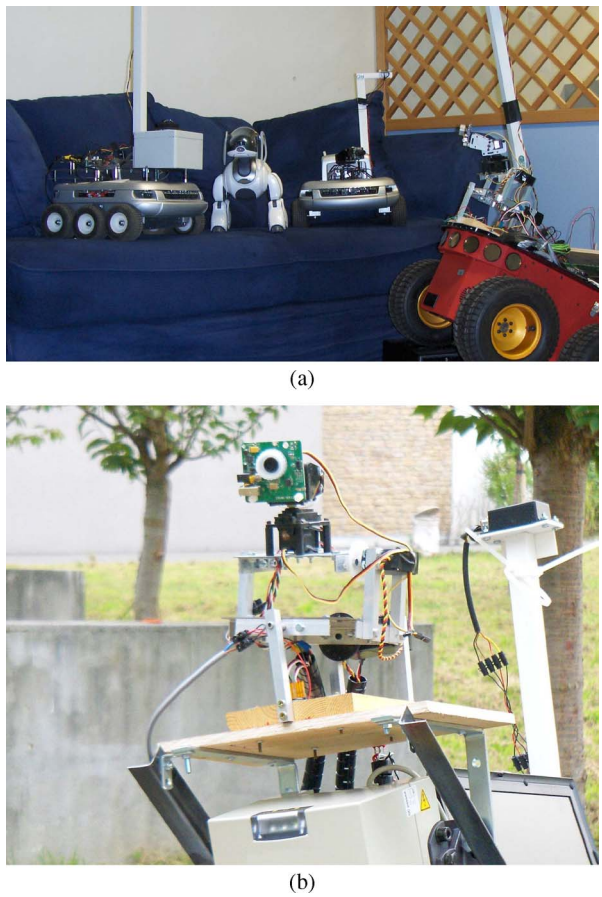

Fig. 4. (a) Wheeled and legged robots used to study bio-inspired navigation. The left robot uses an omnidirectional camera, the right robot uses a FireWire camera mounted on a gyrostabilized pan-tilt platform, and the wheeled robot in the center uses a classical pan-tilt camera. All the robots are provided with a magnetic compass (CMPS03). However, in [60] and [61], we showed that the magnetic compass can be replaced by a visual compass associated with a path integration system. We are also trying to adapt the system to legged robots such as AIBO. (b) Gyrostabilization platform used for experiments on rough terrains.

[47] that can maintain a constant performance level independent of the size of the environment. Several of these limitations can be overcome by using a PerAc Architecture: 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 [54]–[59], but we claim that the PerAc architecture is extremely efficient for spatial behavior learning since it embeds the problem of the environmental partitioning, as well as that of action policy learning.[2] The next section will address the problem of the autonomous building of behavioral attraction basins by HRIs. The problem is treated as a machine learning problem through an interactive demonstration.

We used the following platforms and electronic equipment to study mobile robot navigation [see Fig. 4(a)]:

1) Koala K-Team, pan-tilt camera, magnetic compass;
2) Koala K-Team, omnidirectional camera, magnetic compass;

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

3) Pioneer 2 AT ActivMedia, gyrostabilized platform, pan-tilt camera, magnetic compass.

For outdoor experiments on rough terrains, we built a gyrostabilization platform in order to deal with the effects of a nonplanar surface [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 many reinforcement learning methods [62], [63], such as Q-learning [56], [57], TD(λ) [55], policy gradient reinforcement learning [58], or value and policy search [59]. Our approach differs from these 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 guides the design of our architecture. Our approach also differs because our goal is to design 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 assign a score to each encountered state or state–action event of the environment, corresponding to an expected reward. Reinforcement algorithms that are based on the propagation of rewards converge too slowly in a continuous environment because 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 have also been proposed [64]. As far as HRIs are concerned, we cannot accept a slow acquisition of the behavior (even if sure and optimal). Rather, the behavior must be acquired (and the knowledge must be usable) in a very short time. If an algorithm is allowed to spend time estimating the state space, this time should be used in parallel to the estimation of the topology of the environment. The estimation of the state space topology gives access to a cognitive map which can compute the latent learning of many unrewarded paths [41], [44], [65]–[70]. Evidence of a such latent learning in mammalian species was demonstrated in 1948 by Tolman [71], who showed that the time it takes for a rat to find a goal does not decrease once the reward is found but decreases latently with the number of times the future goal path is experienced before the reward is discovered. Thus, once a reinforcement occurs in a given state, efficient (but suboptimal) strategies are directly available from each visited place.

Moreover, continuous state and action spaces are generally treated as discrete after quantization. What has been encouraging researchers in reinforcement learning is the proof of optimality which already exists for various algorithms, mostly in discrete and nonstochastic state and action spaces [56], [72], [73]. However, convergence toward 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 have continuous state and action spaces [74]. It has also been noted that reinforcement learning algorithms may diverge when a function approximation is used instead of a lookup table [75]. In contrast, our sensory-motor architecture takes into account the continuity of both the state and the action spaces. This paper will show that a continuous action space

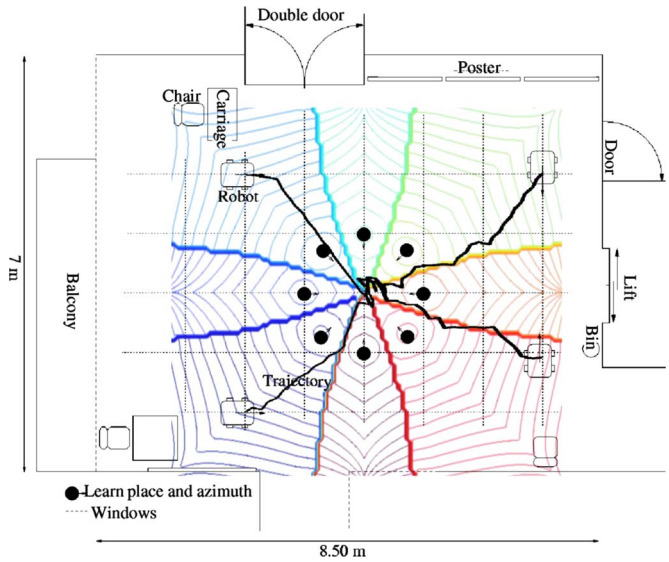

Fig. 5. Real trajectories of homing in an indoor environment with an omnidirectional camera. (Black circles) Eight places are learned at 1 m from the goal (size of the square on the floor). The theoretical place fields are superposed with the map and the trajectories.

enables the measure of an error that aids in the adaptive partitioning of the continuous state space. Moreover, since the suboptimal solutions found in nature for animal navigation are more robust than the current robotics solutions, we can wonder about the need for an optimal algorithm for the learning of spatial behaviors. We can also wonder whether convergence proofs are of interest, considering the time it takes to obtain an efficient suboptimal behavior (with regard to an external measure). Works like [76] have emphasized that reinforcement learning algorithms perform better when they are initialized with a suboptimal policy. The suboptimal solutions computed by our architecture could be used to initialize reinforcement learning algorithms.

Some limitations of classical reinforcement learning algorithms can be overcome by bootstrapping a PerAc architecture (see Fig. 1) [39]. Each place cell is associated with a movement to trigger when the corresponding place is recognized. If the place cells 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 toward the goal from each place in the attraction basin. Learning is equivalent to shaping this basin in order to create an accurate behavioral attractor. Homing or route-following behaviors (see Figs. 5 and 6) can be learned in one shot. Even though human assistance could speed up the convergence [76], 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 framework of an intuitive HRI. In our previous experiments on visual homing or path following (see Figs. 5 and 6), the learning was completely 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

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

well designed for the real-time online learning of skills, in the sense that its goal is to learn associations that occur through direct voluntary experience (concept of enaction [77]). Hence, guiding the robot through the task is more ergonomic than the explicit symbolic communication used in [23] and [78] and should be sufficient for specifying the task to the robot.

In the context of lifelong learning [26], we are presently interested in addressing the problem of the semisupervised building of a behavioral dynamics and its refinement. In addition, 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 visual tracking of the teacher is also possible (closely resembling 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 with the previous place according to the sensory-motor error generated during its traversal.

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 an unknown environment, the environmental partitioning has to be stabilized before 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 shown 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 motor learning neurons A^P (whose elements are a_k^P) is inspired by the Widrow–Hoff (WH) learning rule [79], but other rules are possible.[3] The main difference between a classic WH learning rule and ours is that our rule is composed of two terms: one term for one-shot learning computed as the classic 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 $\epsilon(t) = \sum_{i=1}^{n_P} [p_i^+(t) - p_i^+(t - dt)]^+$, with n_P being the number of place cells, and $[x]^+ = x$ if $x > 0$.

The actions are defined by a population of neurons: Each neuron k in an action group corresponds to a particular orientation $(2 \cdot k \cdot \pi)/n_A$, with 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

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

with $\Delta_k^\theta(t) \in]-\pi, \pi]$ being the shift between the preferred direction $(2 \cdot k \cdot \pi)/n_A$ of the neuron k and the performed movement $\theta(t)$ (here, $\sigma = \pi/6$).

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

$$a_k^M(t) = \epsilon^M \cdot a_k^R(t) + \left[a_k^R(t - dt) - I^R \cdot \epsilon(t)\right]^+ \tag{2}$$

with ϵ^M being a rate guaranteeing that $a_k^M(t)$ will not be greater than one until $1/\epsilon$ steps without reset ($\epsilon^M = 0.001$ for example) and with I^R being a strong positive signal that resets the memory of a_k^M ($I^R > 1/\epsilon^M$ for example).

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

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

with $a_{\max}^M(t) = \max_{k=1,\ldots,n_A}(a_k^M(t))$, used for the normalization (with 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.

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

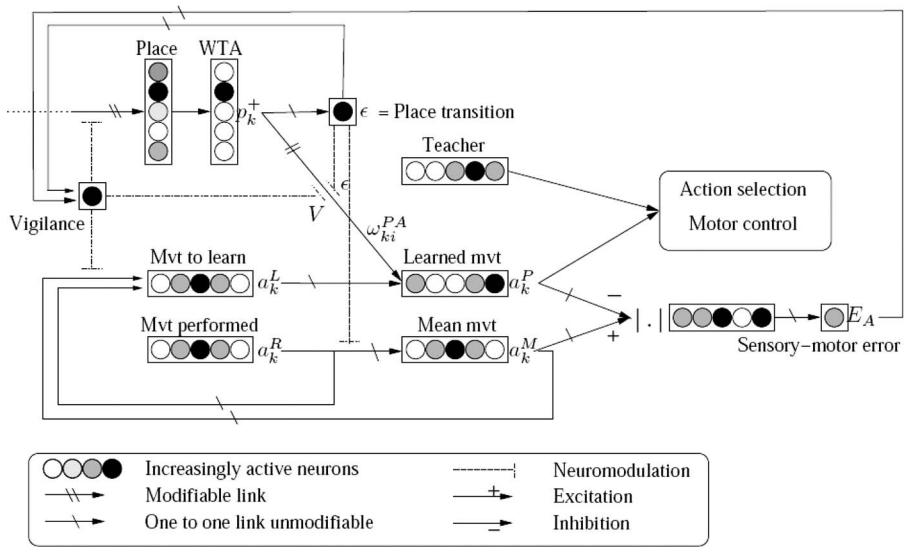

Fig. 7. Modified PerAc architecture enabling either the one-shot learning of places and place–action associations or the refinement of the sensory-motor dynamics. The computation of a signed angular error between the mean performed movement and the predicted movement in a given place allows us to adapt the movement associated with 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.

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

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

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

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, $s_{\max} = \max_{k=1,\ldots,n_A}(s_k)$ 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 and called *Mvt to learn* in Fig. 7). Equation (4) corresponds to the predicted output, and (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 (with 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 mean 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 activity according to the following equations:

$$\frac{d\omega_{ik}^{PA}}{dt} = \left(G_{ik}^i(t) + G_{ik}^d(t - dt) \right) \cdot \epsilon(t) \tag{6}$$

with

$$G_{ik}^i(t) = \left(a_k^L(t) - s_k(t) \right) \cdot p_i^+(t) \cdot V(t) \tag{7}$$

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

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 non-null terms $G_{ik}^i(t)$ in (6). Otherwise, a delayed adaptation is performed each time $\epsilon(t)$ spikes because of the term $G_{ik}^d(t - dt)$ (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 reenters it. 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 sensory-motor learning, this delayed adaptation seems to be crucial for controlling the instants and the contents of the learning.

The remaining question is related to the control of the vigilance signal: What 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 dynamics. The simplest solution for triggering the coding of a new place is to fix a low threshold t_P^- on the place cell activity. If the activity $p^M(t) = \max_{k=1,\ldots,n_P}(p_k(t))$ of the most activated place cell is under this threshold, a new place is learned: $V(t) = \Gamma_0(t_P^- - p^M)$, with Γ_x being the Heaviside function ($\Gamma_x(y) = 1$ if $y > x$, and zero 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 shown in Fig. 8(a) and (b). Since the sampling of the state space controls the precision of the behavioral dynamics, the partitioning of the environment

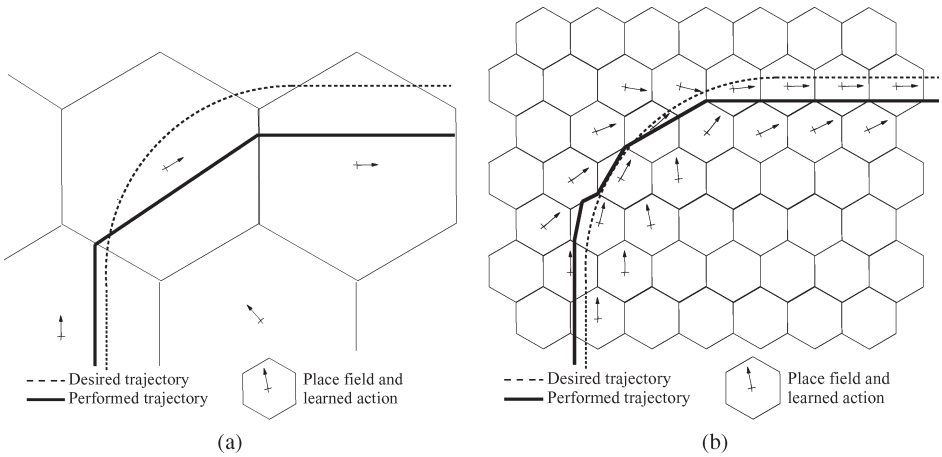

Fig. 8. (a) Example of a regular spatial partitioning with large cells. A movement direction is associated with each cell. The precision of the reproduced trajectory depends on the precision of the state space coding (the size of the cells). (b) Example of a regular spatial partitioning with small cells.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

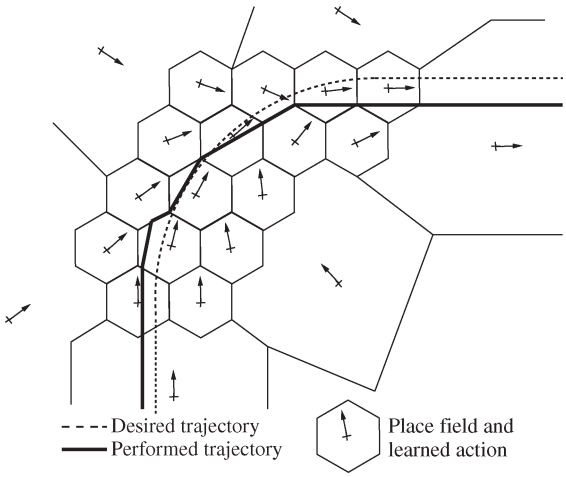

Fig. 9. Example of an adaptive spatial partitioning. The size of the cells is adapted to the complexity of the trajectory. Small cells are used to precisely follow the bend, whereas big cells are used to create a convergence around the trajectory.

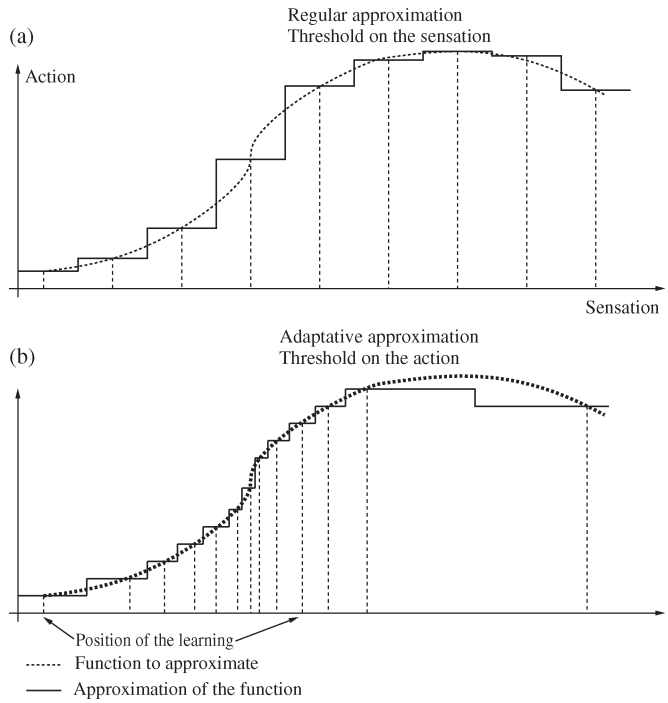

Fig. 10. Illustration of a 1-D example of two approximation methods. (a) Regular partitioning, based on a low threshold on the place cell activity (here, corresponding to a given distance between the position of learning on the "sensation" axis), under which a new place is learned. (b) Adaptive partitioning, 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.

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 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 $\Delta D/\Delta S$ of the sensory-motor function remains valid (the compression factor is adapted to the variation in the 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 in the 1-D example in Fig. 10. With the regular partitioning, the more the function varies, the higher the error is. Yet, when the function varies, the probability of generating diverging trajectories is higher. On the contrary, with the adaptive partitioning, the density of the 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.

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 $\Delta D/\Delta S$. Indeed, the sensory-motor error is higher in complex parts of the trajectory than in easier parts because more changes of the direction occur. Hence, the sensory-motor error appears to be a pertinent signal

for controlling place learning. 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°, then a 90° bend should be encoded by at least three place cells. Thus, this measure can be used to control the vigilance signal in order to adapt the learning location of the place cells to the complexity of the desired behavior. The vigilance signal is defined as

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

In order to avoid overcoding 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 capabilities 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 sufficiently recognized. t^-_P must be correlated with the generalization capabilities.

D. Simulated Environment and Simulated Teacher

In order to theoretically validate our approach, a simulated environment is used. In this environment, the system creates perfect place cells because all the possible landmarks, as well as their identity and their exact azimuth, are provided. In order to simulate human guidance of the robot, an ordered set D of points d_i that parameterizes the desired trajectory is predefined. A dynamic process, where trajectories converge toward 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 $d_{i+\Delta_P}$. $\Delta_P > 0$ depends on the distance[4] d between two points d_i and d_{i+1} ($\Delta_P = 5$ in our simulated environment with $d < 7.5 \times \sqrt{2}$) and on the distance d_r traveled by the agent between each step (here, $d_r < 1$). The dynamic system for human guidance is shown in Fig. 11(a). Fig. 11(b) shows, for a given set D that parameterizes the desired trajectory, the trajectories generated by the dynamic human guidance system. The generated attractor corresponds to the expected robot behavior after learning (i.e., the attractor corresponds to the desired trajectory).

E. Validation of the Proposed System

The experiment in 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 button in order to correct the robot's behavior and teach it the desired trajectory. The strategy of the teacher (deciding when the button should be pressed) is the subject in the next section. Fig. 12(a) and (b) shows 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. Fig. 12(c) also shows the position of the learned places, superimposed with the attractor. The

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

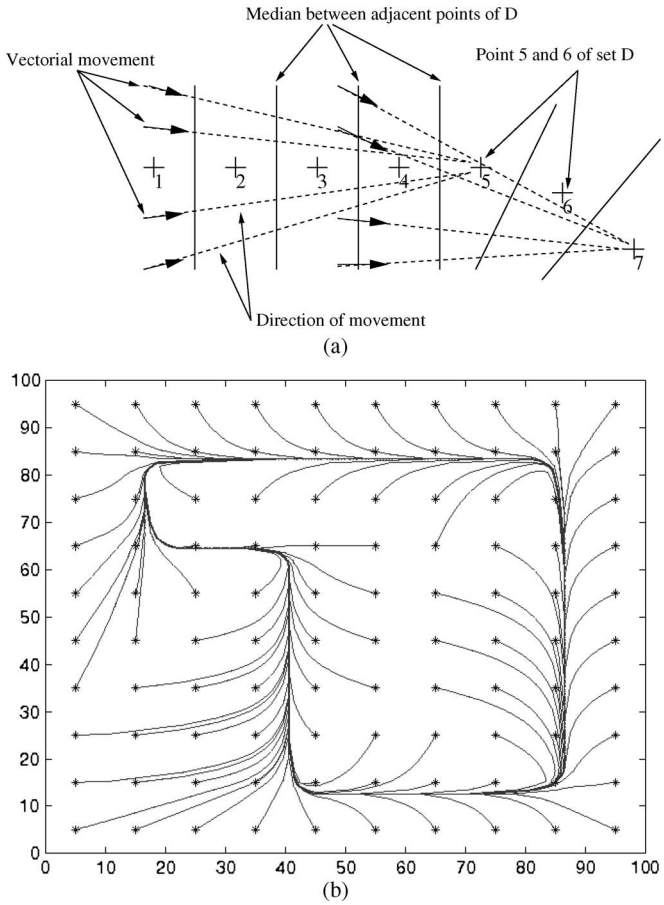

Fig. 11. (a) To simulate human guidance, an ordered set D of points d_i that parameterizes the desired trajectory and a dynamic process, where trajectories converge toward an attractor considered as the optimal trajectory, is used. (b) Trajectories generated by the dynamical process in (a). The trajectories converge toward an attractor defining the optimal trajectory.

simulated robot adapts the density of learned locations to the complexity of the desired trajectory: During bends, 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 such as straight lines.

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

IV. HRIs as a Cognitive Catalyst for the Learning of Behavioral Attractors

In this section, accuracy measures of the reproduced trajectories as compared with the optimal trajectory are proposed. The importance of the interaction loop between the human and the robot during learning, particularly 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 robot's position to be tracked in the Cartesian space. An experiment in an outdoor environment is also proposed.

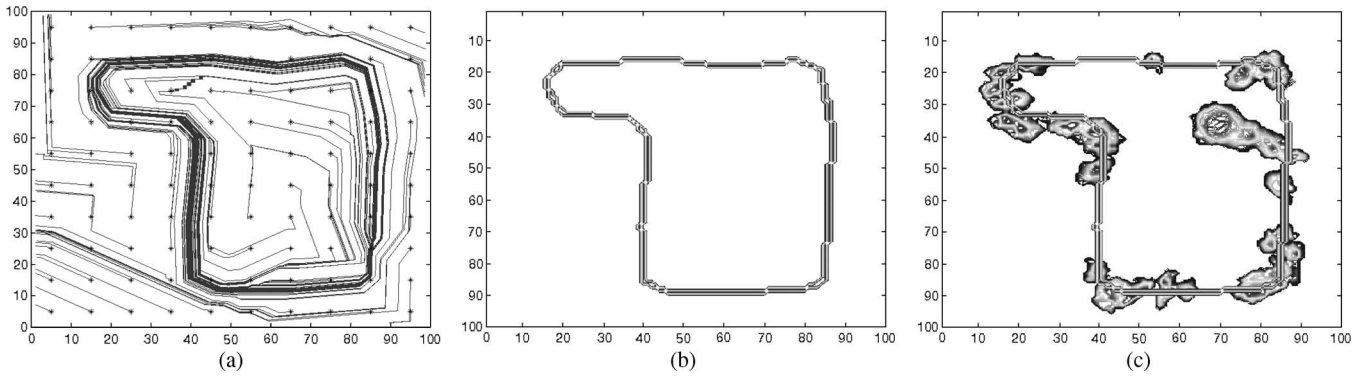

Fig. 12. Training a simulated robot to perform a given trajectory: A human pushes a button to trigger the guidance process defined in Fig. 11. (a) Trajectories generated after the learning. (b) Generated attractor: Mean position of the robot for different starting points in the attraction basin, after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

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 take a very long time to evaluate the complete behavior in the whole environment or to estimate the optimal behavior in each position, we prefer to compare algorithm performance by trying to evaluate the precision of the generated trajectory, from its starting point to its endpoint with respect to a desired trajectory.

Evaluating the spatial precision of a trajectory independently of the temporal variable is an extremely hard problem. Indeed, comparing trajectories without time aspects is equivalent to comparing the sequence of points defining the two trajectories. We propose two measures for comparing the optimal trajectory $\{x_i(p)/p \in \{1, \ldots, P\}\}$ with the reproduced trajectory $\{x_r(t)/t \in [t_i, \ldots, t_f]\}$. The first equation evaluates the mean distance between the robot's position and the closest point of the desired trajectory

$$e_t = \frac{\int_{t=t_i}^{t=t_f} \min_{p=1}^{P} \|x_r(t) - x_i(p)\| \cdot \mathrm{d}t}{t_f - t_i}. \tag{10}$$

This measure is not enough because the robot can navigate very close to the desired trajectory but stay very far from a given point. 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 traveled close enough to each point on the desired trajectory

$$e_p = \frac{\sum_{p=1}^{P} \min_{t=t_i}^{tf} \|x_r(t) - x_i(p)\|}{P}. \tag{11}$$

This second equation is also insufficient because the robot can navigate close to each point of the desired trajectory and then go far off course without increasing the measure e_p. However, the conjunction of both equations allows us to evaluate whether each robot's position was always close to a point on the desired trajectory or the robot traveled close to each point on the desired trajectory. Hence, a combined measure, such as $(e_t + e_p)$, may also be used.

It should be noted that each measure varies in an 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 on the desired trajectory), and e_p should have shrunk to a weak value (the robot has been close to each point on 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 that in the correct direction. Moreover, oscillating around the ideal trajectory may provide 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 before using these measures.

In the following, these two measures will be used in a simulated environment and in a real indoor environment. However, because they require precise measurements of the robot's position, it is far more difficult to use these measures in larger environments. In real outdoor environments, a differential GPS seems necessary. In indoor environments, systems based on a network of calibrated cameras that track the robot across several rooms might be possible. However, we did not have access to these technologies for our experiments. Hence, in the outdoor experiment, the difficulty in estimating the robot's precise position precluded the use of these two measures.

B. Effects of the Interaction Strategy

Our main objective in this section is to demonstrate the importance of HRI and real human guidance as opposed to a

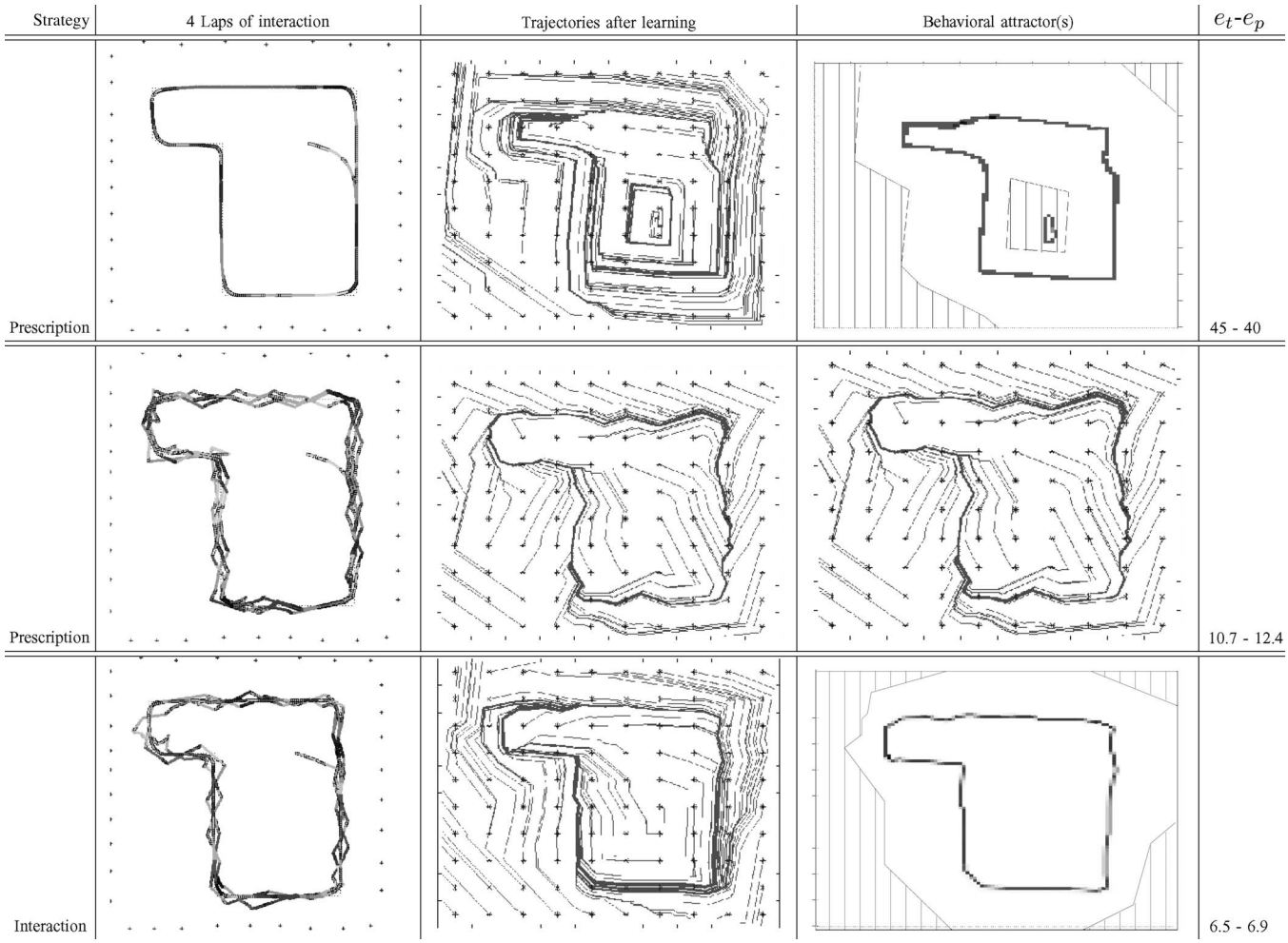

Fig. 13. Left figures show the trajectories during the four laps of training. The figures in the center show some generated trajectories. The behavioral attractors and their attraction basins are displayed on the right figures (the attractors correspond to the mean position of the robot after a long time for different starting points, and the attraction basin is deduced from the figure 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 toward 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 has difficulty stabilizing 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.

predefined strategy such as 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 medium, different strategies may be adopted by the teacher to interact with the robot. The teacher may guide the robot perfectly by adopting a *prescriptive teaching* strategy or, on the contrary, adopt a *proscriptive teaching* approach that consists of correcting the robot when it is too far from the center of the trajectory. These two opposing strategies bring to mind the opposing objectivist and constructivist approaches in robot autonomy, pointed out in [80]. In both cases, the robot should be able to extract the information and use it as well as possible. The result of the experiment shown in 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 must also show the precise trajectory to refine the center of the attraction basin. If the teacher only adopts one of the two strategies, the resulting behavior is expected to be worse 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. These expected results are validated using the same experimental conditions described in the previous section.

2) Experimental Validation: Let us first consider a prescriptive teaching strategy (first line in Fig. 13). As the teacher always performs the same action in the same places without observing the robot's behavior, he never knows whether the robot learns or it is able to reproduce the behavior. Hence, neither the teacher nor the robot knows whether 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

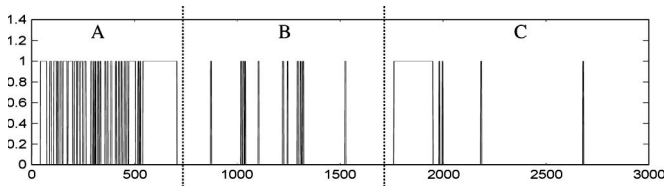

Fig. 14. Rhythm of the interaction between a human teacher and the robot during an experiment like the last one in Fig. 13 (but with the number of training laps not constrained). The graph shows when the human presses the button. The phases of proscription correspond to the Dirac pulses. The phases of prescription correspond to the longer steps. Three periods emerge: During period A, corresponding to the beginning of the interaction, the correction frequency is high—the teacher has to be authoritarian since the robot knows nothing. Period B is characterized by an alternation of correction and observation phases. 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 commands.

center 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 that the teacher can adopt is to correct the robot when it is too far from the center of the trajectory, using a proscriptive teaching strategy (second line in Fig. 13). Hence, the robot oscillates from one border of the allowed path to the other. This strategy has the advantage that the teacher directly evaluates 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 toward its center. The drawback is that the robot does not stabilize on the precise trajectory but oscillates around it. Fig. 13 shows the oscillating effects of using proscriptive teaching alone. This figure also shows that the approximated dynamics no longer has any erroneous parasitic attractor and that the generated attractor is far more accurate (the measures e_t and e_p are divided by four as compared with the result of the prescriptive teaching). The simulations of the prescriptive as well as the proscriptive strategies are, in fact, *ad hoc* processes of guidance that do 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 performs the learned behavior when the button is released), both kinds of learning will naturally emerge from the interaction (see the last line in Fig. 13). During the natural course of the interaction, the teacher oscillates between precise demonstrations of the trajectory (prescriptive teaching), observation of the robot's behavior, and proscriptive corrections as shown in Fig. 14, which shows the rhythm of the interaction (the number of training laps in this experiment was not constrained). The HRI is real, and it takes place by means of a nonverbal nonsymbolic language based on the actions (imposed by the teacher and reproduced by the robot). The teacher alternates between prescriptive and proscriptive phases. As a result, we can see that the generated trajectories are more precise than when proscriptive teaching alone was used. Both strategies have complementary properties and occur successively during the real interaction. The proscriptive teaching strategy enables the creation of the border of the attraction basin, guaranteeing a convergence toward the center of the trajectory, whereas the prescriptive teaching strategy enables the precise digging of the attraction basin center.

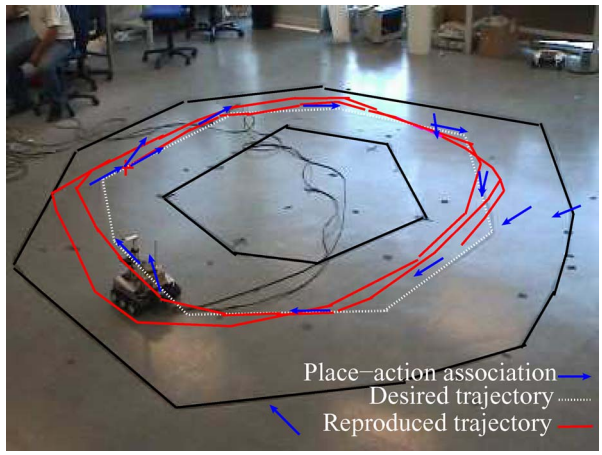

Fig. 15. Indoor experiment: The robot is guided by a human operator. Three laps are sufficient to train the robot to perform the task within the path defined by the black border (not visible to the robot).

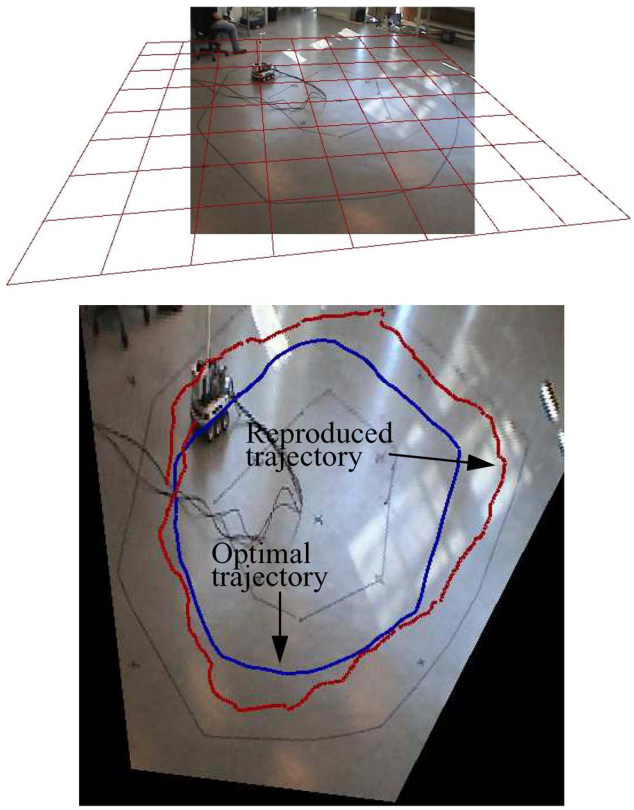

Fig. 16. Measure of an indoor trajectory. The perspective effect of the camera used to record the experiments is first corrected. Then, the user specifies the optimal trajectory. The tracking of the robot in the corrected image allows e_t and e_p to be computed. In this experiment, $e_p = 23$ cm, and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

C. Experiments With Real Robots

We present here the 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 environments. The indoor experiments (see Figs. 15 and 16) demonstrate, in favorable conditions (constant artificial light, horizontal ground, and various and numerous visual landmarks), that it is quite easy to train a robot to perform a sensory-motor task with our system. In the experiment in

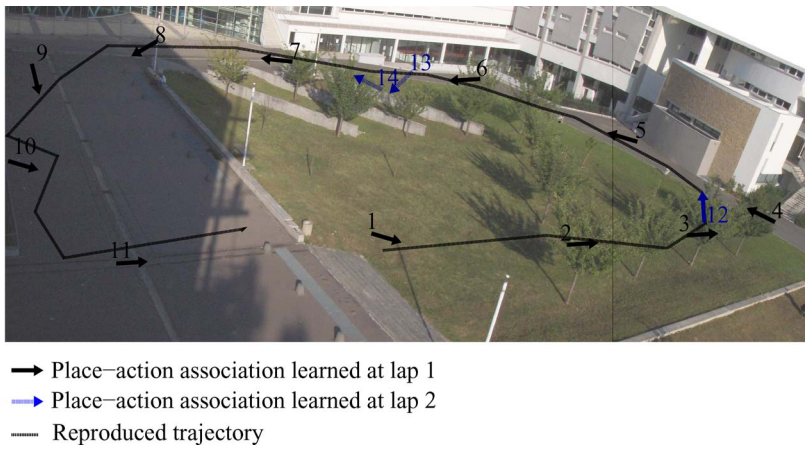

Fig. 17. Outdoor experiment of interactive teaching of a visual path. The 200 m is covered in 9 min. The architecture is split on four processors, but the speed limitation is due to the low level drivers of the robot actuators. Otherwise, the robot's speed could be higher.

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 longer. 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 was used. Fig. 16 shows the tracking and the perspective correction used to measure e_t and e_p. In this experiment, $e_p = 23$ cm, and $e_t = 26$ cm.

Outdoor experiments are far more difficult to analyze due to the constraints of natural and rough environments. Map-building autonomous systems, which are able to deal with outdoor constraints, are rare (see [81] for an example). The size and the nature of the experimental environment prevented us from recording the precise trajectory: Two or three synchronized cameras would have been necessary, and the GPS fails in such *urban canyons*. Moreover, a gyrostabilized platform using two accelerometers was necessary. The camera and the magnetic compass were mounted on this stabilized platform [see Fig. 4(b)] to deal with the nonplanarity of the ground, leading, otherwise, to errors in compass and vision measurements. This platform enables us to limit the effects of a nonplanar 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 in luminosity when the robot's camera captures buildings directly illuminated by the sun as compared to shadowed areas. We had developed an exposure time and gain adaptor to control the parameters 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, within 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. Discussion

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 used directly. Hence, the teacher can immediately see the consequences of his guidance when a place-action association is being learned. The one-shot learning is also instantaneously usable by the robot, giving real feedback to the human on how its actions affect the learning. As a simple one-shot association does not allow the behavioral dynamics to be refined, 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 imprecise guiding. When the robot crosses a location, it integrates the performed movements, regardless of whether they are performed actively or passively (i.e., are decided by the robot or imposed by the teacher). When the robot enters another location, the integrated movement can be used to adapt the learned movement associated with the previous location. Hence, the corrections provided by the teacher enable the dynamics to be refined, whereas the autonomous movements of the robot reinforce the learned dynamics. Such an adaptive learning process enables an approximation of the desired behavior that is as precise as the spatial partitioning and the behavior of the teacher allow.

A crucial problem in the interaction between humans and robots is that the human never knows whether the task is correctly learned by the robot, and the robot never knows whether the teacher is satisfied with its behavior. As neither one can evaluate the other, how is it possible for the robot or for the human to know that the task has been learned? Since the teacher corrects the robot, he cannot know how the robot would have behaved if no correction had been made. Hence, prescriptive teaching alone is insufficient to produce a constructive interaction, as previously illustrated. The teacher has to evaluate the robot's behavior by both prescriptive and proscriptive teaching. Moreover, during such an HRI, a real action-based communication emerges [78]: 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 alternating phases: prescription, proscription, and observation/demonstration. The three phases alternate in time and space as the teaching evolves; these alternating phases define an interaction rhythm. We pointed out (Section IV) that the richness of a real HRI acts as a cognitive catalyst, enhancing the precision of the reproduced behavior, as well as its functioning domain (size of the attraction basin) [5]. This is in contrast to a predetermined guidance strategy such

as prescriptive guidance or a proscriptive strategy consisting of correcting the robot when it commits errors. Prescriptive guidance, commonly referred to as programming by demonstration, is, in fact, far from being an interactive learning process since the robot's behavior does not influence the teacher's. Proscriptive guidance, on the other hand, is the first step toward interactive learning since the robot's behavior does influence the teacher's, which, in turn, influences the robot's and so forth. The system was finally validated by experiments in real indoor and outdoor environments. The experiments prove that the system is mature and could be used in the human world, even by naive operators, in order to teach a patrol mission in an *a priori* 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 or dissatisfied with the robot's behavior. If the teacher is not satisfied with the robot's behavior while the learning has already converged, this must mean that the desired precision is not reachable by the robot, which has performed the task as well as it can. However, even if the teacher is satisfied with the robot's behavior, the robot may nonetheless be aware of 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 could disobey the teacher in known states or express its need for help in unmastered states. Such variations in behavior would constitute another excellent feedback 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 whether its learning enables it to progress or its predictions are standard with respect to the current situation. We are currently working on a progress-based approach derived from [14] and [82] for the metacontrol of the learning, with the goal of giving self-evaluation capabilities to the robots. In [10], [83], and [84], we proposed a progress-based neural architecture which provides the robot with the capability to detect phases of progress, phases of stagnation, and novelty. Novelty detection leads the robot to readapt its erroneous learning.

We also want to investigate how to provide the robot with specific behaviors in response to its self-evaluation of how well it has mastered the task at hand, in order to enrich the interaction and speed up the knowledge transfer. In unmastered situations, the robot could use repair strategies to regain the teacher's attention, by means of a particular behavior (oscillation, stopping, looking toward the teacher, etc.) [85] or by means of a more easily understandable medium such as an expressive robot head [86], [87]. Seeing these behavioral oscillations, the teacher should interact with the robot by giving it the correct orientation, thus providing additional examples for the learning. In mastered states, the robot could exhibit curiosity by choosing not to realize the learned behavior and disobey the teacher in order to find less mastered states in which it can still progress. Communication based on the expression of emotional states could, once again, be very pertinent. Indeed, this could lead the robot toward states that it would not have experienced if it had performed what it had learned or if it had performed the predictable actions imposed by its teacher. Obviously, autoevaluation capabilities also appear to be an excellent starting point to deal with permanent environmental changes or morphological changes in the robot: Self-evaluation capabilities could more easily lead to the development of efficient relearning strategies in the case of permanent changes.

VI. Conclusion and Perspectives

This paper investigated 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 using the sensory-motor error to trigger the learning of new places, 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, despite its simplicity, creates a real HRI with the emergence of an *action-based dialogue*. We have proposed accuracy measures and highlighted the fact that HRIs 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 navigation model, neurons in the hippocampus proper (CA1/CA3 regions) learn and predict transitions between successive multimodal states [44]. A cognitive map performs latent learning of the spatial topology of the environment [71] and can be used to compute a plan of actions to reach an arbitrary goal [68]. The system has been recently validated in a long random exploration experiment (45 min, 3000 steps of the place cell architecture), in a real indoor environment [41]. The experiment highlights the capability of the system to predict place transitions, to latently build the cognitive map of the learned transitions, and thus to plan trajectories to particular goals specified by a simple reinforcement at the goal position at the end of the exploration. The influence of our progress-based metacontroller [83], [84] will be evaluated at every level of this architecture. We will also study how an agent can autonomously detect that it is not really meeting its objective. We will ask how an emotional system could be used as a second-order controller [86] 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 building a single architecture that would allow the robot to deal with spatial as well as temporal modalities (place-action and duration–action strategy), in navigation as well as in robotic arm manipulation [42]. Our goal is to build a merged control architecture for applications in which navigation and object manipulation are considered. A simple example could be a robot that must be able to use door handles or press elevator buttons to achieve its mission. However, such missions also imply the incorporation of object recognition, visual affordance detection [1], [88]–[90], and more sophisticated mechanisms for understanding the natural and/or human world. The adaptation of our system on unmanned aerial vehicles is also currently being studied.

ACKNOWLEDGMENT

The authors would like to thank G. Désilles for the investment in our research, J. P. Banquet for the biological counterpart of the model, and all the members of the Neurocybernetic Team for the daily support and the animated discussions on cognitive robotics and more. Movies of the experiments presented in Figs. 6, 15, and 17 are available on the authors' website and at http://www.etis.ensea.fr/~neurocyber/giovannangeli/Home.html.

REFERENCES

[1] J. Gibson, *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin, 1979.
[2] F. Varela, *Invitation aux Sciences Cognitives*. Paris, France: Seuil, 1996.
[3] J. von Uexküll, *Mondes Animaux et Monde Humain*. Paris, France: Gonthier, 1966.
[4] V. Klingspor, J. Demiris, and M. Kaiser, "Human–robot-communication and machine learning," *Appl. Artif. Intell.*, vol. 11, no. 7/8, pp. 719–746, 1997.
[5] C. Giovannangeli and P. Gaussier, "Human–robot interactions as a cognitive catalyst for the learning of behavioral attractors," in *Proc. 16th IEEE Int. Symp. Robot Human Interactive Commun.*, Jeju, Korea, 2007, pp. 1028–1033.
[6] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robot. Auton. Syst.*, vol. 42, no. 3/4, pp. 143–166, Mar. 2003.
[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," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 35, no. 4, pp. 471–480, Jul. 2005.
[8] H. C.-H. Hsu and A. Liu, "A flexible architecture for navigation control of a mobile robot," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 3, pp. 310–318, May 2007.
[9] S. Thrun, "Robotic mapping: A survey," in *Exploring Artificial Intelligence in the New Millenium*, G. Lakemeyer and B. Nebel, Eds. San Francisco, CA: Morgan Kaufmann, 2002.
[10] C. Giovannangeli, "Navigation autonome en environnement intérieur et extérieur: Apprentissage sensori-moteur et planification dans un cadre interactif," Ph.D. dissertation, Université de Cergy-Pontoise, Cergy-Pontoise, France, 2007.
[11] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, "Real-time visual loop-closure detection," in *Proc. IEEE ICRA*, 2008, pp. 1842–1847.
[12] J. Schmidhuber, "Curious model-building control systems," in *Proc. Int. Conf. Neural Netw.*, Piscataway, NJ, 1991, vol. 2, pp. 1458–1463.
[13] J. Schmidhuber, J. Zhao, and M. Wiering, "Reinforcement learning with self-modifying policies," in *Learning to Learn*. Norwell, MA: Kluwer, 1997, pp. 293–309.
[14] F. Kaplan and P.-Y. Oudeyer, *Maximizing Learning Progress: An Internal Reward System for Development*, F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, Eds. New York: Springer-Verlag, 2004.
[15] P.-Y. Oudeyer, F. Kaplan, V. Hafner, and A. Whyte, "The playground experiment: Task-independent development of a curios robot," in *Proc. AAAI Spring Symp. Workshop Dev. Robot.*, D. Bank and L. Meeden, Eds., 2005, pp. 42–47.
[16] C. Tovey and S. Koenig, "Improved analysis of greedy mapping," in *Proc. IEEE Int. Conf. IROS*, 2003, pp. 3251–3257.
[17] H. Andreasson, A. Treptow, and T. Duckett, "Localization for mobile robots using panoramic vision, local features and particle filter," in *Proc. IEEE ICRA*, Barcelona, Spain, 2005, pp. 3348–3353.
[18] P. Gaussier, C. Joulain, J. Banquet, S. Leprêtre, and A. Revel, "The visual homing problem: An example of robotics/biology cross fertilization," *Robot. Auton. Syst.*, vol. 30, no. 1/2, pp. 155–180, Jan. 2000.
[19] C. Giovannangeli, P. Gaussier, and G. Désilles, "Robust mapless outdoor vision-based navigation," in *Proc. IEEE/RSJ Int. Conf. IROS*, Beijing, China, 2006, pp. 3293–3300.
[20] P. Andry, P. Gaussier, S. Moga, J. Banquet, and J. Nadel, "Learning and communication in imitation: An autonomous robot perspective," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 31, no. 5, pp. 431–442, Sep. 2001.
[21] 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
[22] M. Nicolescu and M. Mataric, "Learning and interacting in human–robot domains," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 31, no. 5, pp. 419–430, Sep. 2001.
[23] T. W. Fong, C. Thorpe, and C. Baur, "Robot, asker of questions," *Robot. Auton. Syst.*, vol. 42, no. 3/4, pp. 235–243, Mar. 2003.
[24] J.-H. Hong, Y.-S. Song, and S.-B. Cho, "Mixed-initiative human–robot interaction using hierarchical Bayesian networks," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 37, no. 6, pp. 1158–1164, Nov. 2007.
[25] A. Poncela, C. Urdiales, E. Perez, and F. Sandoval, "A new efficiency-weighted strategy for continuous human/robot cooperation in navigation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 3, pp. 486–500, May 2009.
[26] S. Thrun and T. Mitchell, "Lifelong robot learning," *Robot. Auton. Syst.*, vol. 15, no. 1, pp. 25–46, Jul. 1995.
[27] G. Hayes and J. Demiris, "A robot controller using learning by imitation," in *Proc. 2nd Int. Symp. Intell. Robot. Syst.*, Grenoble, France, 1994, pp. 198–204.
[28] K. Dautenhahn, "Getting to know each other: Artificial social intelligence for autonomous robots," *Robot. Auton. Syst.*, vol. 16, no. 2–4, pp. 333–356, Dec. 1995.
[29] C. G. Atkeson and S. Schaal, "Robot learning from demonstration," in *Proc. 14th ICML*, 1997, pp. 12–20.
[30] P. Gaussier, S. Moga, M. Quoy, and J. Banquet, "From perception–action loops to imitation processes: A bottom–up approach of learning by imitation," *Appl. Artif. Intell.*, vol. 12, no. 7/8, pp. 701–727, Oct.–Dec. 1998.
[31] A. Billard and M. J. Mataric, "Learning human arm movements by imitation: Evaluation of a biologically inspired connectionist architecture," *Robot. Auton. Syst.*, vol. 37, no. 2/3, pp. 145–160, Nov. 2001.
[32] A. Alissandrakis, C. Nehaniv, and K. Dautenhahn, "Imitation with ALICE: Learning to imitate corresponding actions across dissimilar embodiments," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 32, no. 4, pp. 482–496, Jul. 2002.
[33] P. Andry, P. Gaussier, J. Nadel, and B. Hirsbrunner, "Learning invariant sensorimotor behaviors: A developmental approach to imitation mechanisms," *Adapt. Behav.*, vol. 12, no. 2, pp. 117–138, Oct. 2004.
[34] 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*, K. Dautenhahn and C. Nehaniv, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[35] A. Billard, Y. Epars, S. Calinon, S. Schaal, and G. Cheng, "Discovering optimal imitation strategies," *Robot. Auton. Syst.*, vol. 47, no. 2/3, pp. 69–77, Jun. 2004.
[36] S. Calinon, F. Guenter, and A. Billard, "On learning, representing and generalizing a task in a humanoid robot," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 286–298, Apr. 2007.
[37] S. Calinon and A. Billard, "Active teaching in robot programming by demonstration," in *Proc. IEEE Int. Symp. RO-MAN*, 2007, pp. 702–707.
[38] S. Calinon and A. Billard, "What is the teacher's role in robot programming by demonstration? Toward benchmarks for improved learning," *Interact. Stud.*, vol. 8, no. 3, pp. 441–464, 2007.
[39] P. Gaussier and S. Zrehen, "PerAc: A neural architecture to control artificial animals," *Robot. Auton. Syst.*, vol. 16, no. 2–4, pp. 291–320, Dec. 1995.
[40] P. Gaussier, A. Revel, C. Joulain, and S. Zrehen, "Living in a partially structured environment: How to bypass the limitation of classical reinforcement techniques," *Robot. Auton. Syst.*, vol. 20, no. 2, pp. 225–250, Jun. 1997.
[41] 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. IEEE/RSJ Int. Conf. IROS*, Nice, France, 2008, pp. 676–683.
[42] M. Lagarde, P. Andry, P. Gaussier, and C. Giovannangeli, "Learning new behaviors: Toward a control architecture merging spatial and temporal modalities," in *Proc. Workshop Interactive Robot Learn.—Int. Conf. RSS*, 2008. [Online]. Available: http://www.dfki.de/cosy/www/events/InteractiveRobotLearning2008/proceedings.php
[43] C. Giovannangeli, P. Gaussier, and J.-P. Banquet, "Robustness of visual place cells in dynamic indoor and outdoor environment," *Int. J. Adv. Robot. Syst.*, vol. 3, no. 2, pp. 115–124, Jun. 2006.
[44] 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," *Biol. Cybern.*, vol. 86, no. 1, pp. 15–28, Jan. 2002.
[45] J. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness," *Behav. Brain Sci.*, vol. 24, no. 5, pp. 939–1031, Oct. 2001.
[46] B. Cartwright and T. Collett, "Landmark learning in bees," *J. Comput. Physiol.*, vol. 151, no. 4, pp. 521–543, Dec. 1983.

[47] D. Lambrinos, R. Moller, T. Labhart, R. Pfeifer, and R. Wehner, "A mobile robot employing insect strategies for navigation," *Robot. Auton. Syst.*, vol. 30, no. 1, pp. 39–64, Jan. 2000.
[48] V. V. Hafner, "Learning places in newly explored environment," in *Proc. SAB*, 2000, pp. 111–120.
[49] A. Argyros, K. Bekris, S. Orphanoudakis, and L. Kavraki, "Robot homing by exploiting panoramic vision," *Auton. Robots*, vol. 19, no. 1, pp. 7–25, Jul. 2005.
[50] V. V. Hafner, "Cognitive maps in rats and robots," *Adapt. Behav.*, vol. 13, no. 2, pp. 87–96, Jun. 2005.
[51] J. Saez Pons, W. Hübner, H. Dahmen, and H. A. Mallot, "Vision-based robot homing in dynamic environments," in *Proc. 13th IASTED Int. Conf. Robot. Appl.*, 2007, pp. 293–298.
[52] L. Smith, A. Philippides, and P. Husbands, "Navigation in large-scale environments using an augmented model of visual homing," in *Proc. 9th Int. Conf. SAB—From Animals to Animats 9*, J. G. Carbonell and E. J. Siekmann, Eds., 2006, pp. 251–262.
[53] L. Smith, A. Philippides, P. Graham, B. Baddeley, and P. Husbands, "Linked local navigation for visual route guidance" *Adapt. Behav.*, vol. 15, no. 3, pp. 257–271, Sep. 2007.
[54] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.
[55] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Mach. Learn.*, vol. 3, no. 1, pp. 9–44, Aug. 1988.
[56] C. Watkins, "Learning with delayed rewards," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K., 1989.
[57] C. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, no. 3/4, pp. 279–292, 1992.
[58] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, no. 3/4, pp. 229–256, May 1992.
[59] L. Baird, "Gradient descent for general reinforcement learning," in *Proc. Neural Inf. Process. Syst.*, 1998, vol. 11, pp. 968–974.
[60] S. Leprêtre, P. Gaussier, and J. Cocquerez, "From navigation to active object recognition," in *Proc. 6th Int. Conf. SAB*, Paris, France, 2000, pp. 266–275.
[61] C. Giovannangeli and P. Gaussier, "Orientation system in robots: Merging allothetic and idiothetic estimations," in *Proc. 13th Int. Conf. Adv. Robot.*, Jeju, Korea, 2007, pp. 349–354.
[62] L. P. Kaelbling, M. L. Littman, and A. P. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, pp. 237–285, 1996.
[63] E. Zalama, J. Gomez, M. Paul, and J. Peran, "Adaptive behavior navigation of a mobile robot," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 32, no. 1, pp. 160–169, Jan. 2002.
[64] R. Munos and J. Patinel, "Reinforcement learning with dynamic covering of state–action space: Partitioning Q-learning," in *Proc. 3rd Int. Conf. SAB—From Animals to Animats 3*, 1994, pp. 354–363.
[65] M. Arbib and I. Lieblich, "Motivational learning of spatial behavior," in *Systems Neuroscience*. New York: Academic, 1977, pp. 77–87.
[66] M. Mataric, "Integration of representation into goal-driven behavior-based robot," *IEEE Trans. Robot. Autom.*, vol. 8, no. 3, pp. 304–312, Jun. 1992.
[67] B. Schölkopf and H. A. Mallot, "View-based cognitive mapping and path-finding," *Adapt. Behav.*, vol. 3, no. 3, pp. 311–348, 1995.
[68] A. Revel, P. Gaussier, S. Leprêtre, and J. Banquet, "Planification versus sensory-motor conditioning: What are the issues?" in *Proc. SAB—From Animals to Animats 5*, 1998, pp. 129–138.
[69] N. A. Schmajuk and H. Voicu, "Exploration, navigation and cognitive mapping," *Adapt. Behav.*, vol. 8, no. 3, pp. 207–223, 2000.
[70] H. Andreasson and T. Duckett, "Topological localization for mobile robots using omni-directional vision and local features," in *Proc. 5th IFAC Symp. IAV*, Lisbon, Portugal, 2004.
[71] E. Tolman, "Cognitive maps in rats and men," *Psychol. Rev.*, vol. 55, no. 4, pp. 189–208, Jul. 1948.
[72] J. N. Tsitsiklis and B. Van Roy, "An analysis of temporal-difference learning with function approximation," *IEEE Trans. Autom. Control*, vol. 42, no. 5, pp. 674–690, May 1997.
[73] S. P. Singh, T. Jaakkola, M. L. Littman, and C. Szepesvári, "Convergence results for single-step on-policy reinforcement-learning algorithms," *Mach. Learn.*, vol. 38, no. 3, pp. 287–308, Mar. 2000.
[74] S. J. Bradtke, "Reinforcement learning applied to linear quadratic regulation," in *Proc. Adv. Neural Inf. Process. Syst.*, 1993, vol. 11, pp. 295–302.
[75] S. Thrun and A. Schwartz, "Issues in using function approximation for reinforcement learning," in *Proc. Connectionist Models Summer School*, M. Mozer, P. Smolensky, D. Touretzky, J. Elman and A. Weigend, Eds., 1993. [Online]. Available: http://www.ri.cmu.edu/publication_view.html?pub_id=672
[76] W. D. Smart and L. P. Kaelbling, "Practical reinforcement learning in continuous spaces," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 903–910.
[77] F. Varela, E. Thompson, and E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press, 1991.
[78] N. Koenig and M. J. Mataric, "Behavior-based segmentation of demonstrated task," in *Proc. IEEE ICDL*, 2006.
[79] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *Proc. IRE WESCON Conv. Rec.*, 1960, vol. 4, pp. 96–104.
[80] J. Stewart, "The implication for understanding high-level cognition of a grounding in elementary adaptive systems," *Robot. Auton. Syst.*, vol. 16, no. 2–4, pp. 107–116, Dec. 1995.
[81] C. Castejon, D. Blanco, and L. Moreno, "Compact modeling technique for outdoor navigation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 1, pp. 9–24, Jan. 2008.
[82] P.-Y. Oudeyer, "Intelligent adaptive curiosity: A source of self-development," in *Proc. 4th Int. Workshop Epigenetic Robot.: Modeling Cogn. Develop. Robot. Syst.*, L. Berthouze, H. Kozima, C. Prince, G. Sandini, G. Stojanov, G. Metta, and C. Balkenius, Eds., 2004, vol. 117, pp. 127–130.
[83] C. Giovannangeli and P. Gaussier, "Learning to navigate, progress and self-evaluation," in *Proc. 6th Int. Conf. Epigenetic Robot.: Modeling Cogn. Develop. Robot. Syst.*, 2006. [Online]. Available: http://www.csl.sony.fr/epirob2006/technicalProgram.htm
[84] C. Giovannangeli, S. Boucenna, and P.Gaussier, "About the constructivist role of self-evaluation for interactive learnings and self-development," in *Proc. IEEE/RSJ Int. Conf. IROS Workshop: From Motor Interact. Learn. Robots*, 2008. [Online]. Available: http://webia.lip6.fr/~sigaud/IROS2008workshop.html
[85] C. Muhl and Y. Nagai, "Does disturbance discourage people from communicating with a robot?" in *Proc. 16th IEEE Int. Symp. RO-MAN*, 2007, pp. 1137–1142.
[86] L. Canamero and P. Gaussier, *Emotion Understanding: Robots as Tools and Models*, J. Nadel and D. Muir, Eds. New York: Oxford Univ. Press, 2005, pp. 235–258.
[87] M. Ogino, A. Watanabe, and M. Asada, "Mapping from facial expression to internal state based on intuitive parenting," in *Proc. 6th Int. Workshop Epigenetic Robot.*, 2006, pp. 182–183.
[88] M. Steedman, "Formalizing affordance," in *Proc. 24th Annu. Meeting Cogn. Sci. Soc.*, 2002, pp. 834–839.
[89] 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," *Adapt.Behav.*, vol. 15, no. 4, pp. 447–472, Dec. 2007.
[90] E. Rome, L. Paletta, E. Sahin, G. Dorffner, J. Hertzberg, R. Breithaupt, G. Fritz, J. Irran, F. Kintzler, C. Lörken, S. May, and E. Ugur, "The MACS project: An approach to affordance-inspired robot control," in *Towards Affordance-Based Robot Control*, vol. 4760, *Lecture Notes in Computer Sciences*. Berlin, Germany: Springer-Verlag, 2008, pp. 173–210.

Christophe Giovannangeli received the Ph.D. degree from Cergy-Pontoise University, Cergy-Pontoise, France, in 2007.

He is currently a Postdoctoral Researcher with the Neurocybernetic Team, Image and Signal Processing (ETIS) Laboratory, Cergy-Pontoise University. He has been a Postdoctoral Researcher with the Computer Science Department, Technion, Israel Institute of Technology, Haifa, Israel. His current research interests are pursuit-evasion games in the presence of obstacles for autonomous UGV and UAV, landmark selection and recognition, online environment learning, action planning, teaching robot behavior through human–robot interactions, and bio-inspired robotics. More generally, he is interested in improving the operational autonomy of robotics systems and control architectures.

Philippe Gaussier is currently a Professor with Cergy-Pontoise University, Cergy-Pontoise, France, where he leads the Neurocybernetic Team of the Image and Signal Processing (ETIS) Laboratory. His current research interests are the modeling of cognitive mechanisms and brain structures such as the hippocampus and its relation to cortical structures such as parietal, temporal, and prefrontal areas; the dynamics of visual perception; and the development of interaction capabilities (imitation, emotions, etc.). His current robotic applications include autonomous and online learning for motivated visual navigation, object manipulation, place learning, visual homing, object discrimination, and action planning. Pr. Gaussier is a member of the Institut Universitaire de France, Paris, France.