



**HAL**  
open science

# Biologically inspired neural networks for spatio-temporal planning in robotic navigation tasks

Julien Hirel, Philippe Gaussier, Mathias Quoy

► **To cite this version:**

Julien Hirel, Philippe Gaussier, Mathias Quoy. Biologically inspired neural networks for spatio-temporal planning in robotic navigation tasks. Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on, Dec 2011, Thailand. pp.1627-1632, <10.1109/ROBIO.2011.6181522>. <hal-00694643>

**HAL Id: hal-00694643**

**<https://hal.science/hal-00694643v1>**

Submitted on 5 May 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Biologically inspired neural networks for spatio-temporal planning in robotic navigation tasks

Julien Hirel, Philippe Gaussier and Mathias Quoy

Final draft version, accepted for publication as:

J. Hirel, P. Gaussier, and M. Quoy, “Biologically inspired neural networks for spatio-temporal planning in robotic navigation tasks”, in *Robotics and Biomimetics (ROBIO)*, 2011 IEEE International Conference on, dec. 2011, pp. 1627-1632. <http://dx.doi.org/10.1109/ROBIO.2011.6181522>

The original publication is available at <http://ieeexplore.ieee.org>

**Abstract**—In this paper we present a biologically-inspired model of spatio-temporal learning in the hippocampus and prefrontal cortex which can be used in tasks requiring the behavior of the robot to be constrained by sensory and temporal information. In this model chains of sensory events are learned and associated with motor actions. The temporality of these sequences is also learned and can be used to predict the timing of upcoming events. The neural network acts as a novelty detector and can modulate the behavior of the robot in case its actions do not have the expected consequences. The system is used to solve two different robotic navigation tasks involving an alternation between random exploration, goal-directed navigation and waiting periods of various lengths.

## I. INTRODUCTION

The hippocampus is an essential part of the brain when it comes to spatial navigation. Its role in navigation has been brought to light with the discovery of *place cells* [1], neurons firing at specific locations in the environment. More recently, some results have suggested that the hippocampus encodes not only places but also accessible paths [2]. One particular experimental protocol has allowed to bring to light the spatio-temporal role of the hippocampus in goal-oriented navigation. Solving this task for rats is done in 3 steps: i) navigate toward a goal zone ii) wait without moving for 2 seconds before a food pellet is dropped from the ceiling iii) explore the area to find the food. Physiological recordings have provided evidence of spatial but also temporal information about the timing of the waiting period being present at the level of the hippocampus [3]. Burton et al. [4] suggest that this activity propagates to the prefrontal cortex and is necessary for the animal to solve the task with good performances. The hippocampal spatio-temporal activity seems to be a key component in the time

estimation system allowing the animal to predict the delayed consequences of its own actions.

The *place preference task* also presents some interesting challenges for robotic navigation. It requires goal-oriented navigation skills, a time-estimation mechanism linked with action selection and the ability to alternate between different behavioral strategies (move toward a known location, wait, explore) depending on the phase of the experiment. The time estimation is crucial for the animal to be able to resume its motion at the end of the waiting period, especially in the absence of the rewarding stimulus. To do so, some knowledge of the length of time spent waiting before obtaining the food must be acquired. The animal must then be able to adapt its behavior depending on the correspondence between its temporal predictions and what it observes. Integrating a temporal aspect is crucial when planning trajectories in a dynamic environment, especially for multi-robot systems where each agent must plan to avoid static obstacles as well as other robots [5]. Predicting when to expect sensory events and detecting their absence could help a robot know when to “replan”. The learned temporal characteristics of the task could thus be used to avoid the computational cost of constant replanning or the need for user-defined heuristics to control when to replan [6], [7].

In an attempt to have a better understanding of the mechanisms involved in spatio-temporal processing in the mammal brain and to apply this knowledge to robotic applications, we developed a biologically-inspired neural network model. We discuss how such a neural architecture can give a robot the ability to control its behavior depending on the timing of perceived sensory events. The robot can learn temporal predictions about the delayed consequences of its actions. As a result, unexpected consequences can be detected and the behavior of the robot adapted thereof. We show how we can use our model to solve the place preference task with a mobile robot. We discuss how this system can be generalized to perform well in any task requiring both navigation and time-estimation skills. An example is given by training the robot to perform a patrol with several checkpoints where it must stop for various periods of time.

## II. MODEL

Banquet et al. [8] presented a theoretical model which predicted the need for transition cells in order to plan with a cognitive map. Transition cells were capable of learning temporal relations between past and present sensory signals. This learning consequently allowed transition cells to predict future sensory events. In [9], sequences of motor actions

This work was supported by the CRNS, DGA and ANR.

J. Hirel, P. Gaussier and M. Quoy are with the ETIS Laboratory, CNRS - ENSEA - University of Cergy-Pontoise, F95000

Corresponding author is [gaussier@ensea.fr](mailto:gaussier@ensea.fr)

were learned and reproduced by learning the timing between motor commands. In [10], transitions between visual place cells were learned and used in navigation with a planning strategy based on a cognitive map of transitions. However no timing was learned for the spatial transitions between places. The cognitive map provided information about the available paths between spatial states but not about the time needed to shift from one state to the next. In an attempt to obtain a unified model for different robotic applications, a learning equation (1) based on a Normalized Least Mean Square (NLMS) algorithm was developed. It allows the model to learn and predict various types of upcoming signals, coming from several modalities (visual and proprioceptive) [11].

Fig. 1 shows the neural network for the learning of the associations. Multi-modal state information is computed from sensory input coming from visual, auditory, vestibular systems etc. A *winner-take-all* (WTA) competition takes place between the state neurons. Any change in the winning state triggers the spectral timing memory. Each state has a corresponding battery of neurons in the memory. These neurons are activated with different timings, forming a spectral decomposition of time [12]. Only one battery can be activated at any time, and its activity gives information about the time passed since the last sensory event (i.e. a change in the sensory state competition). When a new sensory event occurs, the association of this new information (coming from the state neurons) with the spectral timing memory of the last sensory event is learned by state prediction cells. As a result, the system learns to connect the first event with the second and memorizes the timing between the two events. Later when the first event is experienced again, it activates its corresponding spectral timing neurons and predictions for the second event will arise from transitions cells. The shape of the prediction activity looks like a Gaussian curve which reaches its maximum value just before the expected timing of the next event. The equation used for learning the sensory association shown in fig. 1 is the following:

$$\frac{dW_{ij}(t)}{dt} = \alpha \cdot \eta_i(t) \cdot \frac{(X_i^{US}(t) - X_i(t)) \cdot X_j^{CS}(t)}{\sum_k X_k^{CS}(t)^2 + \sigma_1} \quad (1)$$

$$\eta_i(t) = |X_i^{US}(t) - m_i^{US}(t)| + \sigma_2 \quad (2)$$

$$m_i^{US}(t) = \gamma \cdot m_i^{US}(t - dt) + (1 - \gamma) \cdot X_i^{US}(t) \quad (3)$$

$X_i$  is the activity of neuron  $i$ ,  $\alpha$  is the learning rate,  $\eta_i$  a learning modulation,  $X_i^{US}$  is the unconditional signal for the LMS.  $\sigma_1$  is a small value used to avoid the divergence of the synaptic weights for very low memory values.  $m_i^{US}$  is a sliding mean of  $X_i^{US}$ ,  $\sigma_2$  is a low value setting a minimal learning rate and  $\gamma$  a parameter controlling the balance between past and current activities in the computation of the sliding mean.

Sensory events can include the recognition of a new visual place cell, the detection of a color on the floor signaling an important place, the sound being made by the release of the food reward or a change in the proprioceptive information coming from odometry. At any time transition cells can predict reachable sensory states, depending on the current state. The sensory transitions can then become motor

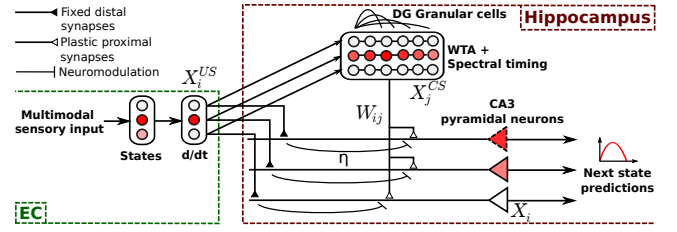


Fig. 1. Hippocampal associative learning architecture

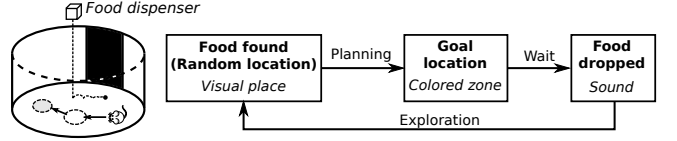


Fig. 2. The place task as a sequence of behavioral strategies cued on multi-modal perceptive events.

transitions when associated with motor actions. Here, the motor actions of the robot can be controlled in terms of orientation and speed, allowing it to control whether or not it is moving. The *place preference task* can then be characterized by a series of behaviors controlled by sensory events (fig. 2). The prediction of the perception of the sound signaling the reward has to be associated with the immobility of the robot, since the reward will only be delivered after a certain period of time spent on the goal location. Other purely spatial predictions are associated with directions to take with a particular speed (a default moving speed). Motor actions associated with each transition are computed from path integration, using the odometry of the robot.

A navigation strategy (e.g. planning with a cognitive map) can use the information about available transitions and bias the selection of a transition and its associated action [10]. The cognitive map forms a topological representation of the environment by linking subsequent transitions. In the map, goals are represented by rewarding transitions (e.g. arriving at a food or water source etc.). Also, each received reward (e.g. food) has an associated drive (e.g. hunger) controlling the tendency of the robot to look for that reward. The reward signal satisfying a particular drive triggers the association in the cognitive map between this drive and the last transition performed. The drive activity can then be propagated through the map (with synaptic weights  $< 1$ ) using a max operator (in a fashion similar to the Bellman-Ford algorithm). The gradient of activity in the map gives information about the distance to rewarding transitions.

Another pathway associates transition activity with satisfaction signals (fig. 3). The predictions emitted by these neurons correspond to actions which should lead to a satisfying state. The equations for the learning of the association (4) and the satisfaction prediction activity (5) are:

$$\frac{dW_{ij}^{TP}(t)}{dt} = \alpha \cdot (1 - W_{ij}^{TP}(t)) \cdot X_j^T(t) \cdot X_j^S(t) - \gamma \cdot X_j^T(t) \quad (4)$$

$$X_i^P(t) = \sum_k W_{ik}^{TS}(t) \cdot X_k^T(t) - \theta \quad (5)$$

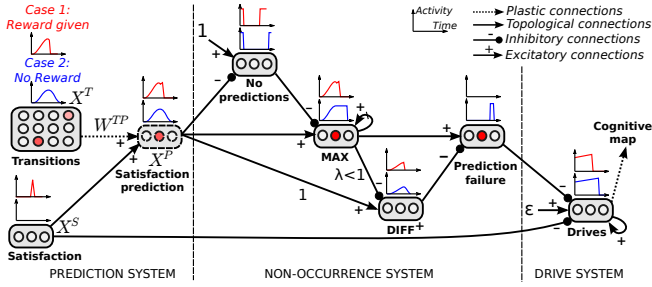


Fig. 3. Detail of the prediction, non-occurrence and drive systems. Transition activity is associated with satisfaction signals. Transitions predicting satisfying stimuli and neutral stimuli can thus be differentiated. A neural network is used to memorize the maximal activity of the bell-shaped predictions. When the prediction activity falls back under a fraction  $\lambda$  of this maximum, it means that the expected timing of the satisfaction has passed and it will probably not happen. The parameter  $\lambda$  controls the tendency of the system to wait more or less for expected satisfactions after their predicted timing. The closer  $\lambda$  is to 1, the faster the system will signal the non-occurrence of a prediction. As long as the prediction activity is close to the max, the non-occurrence signal is inhibited but when the difference becomes significant non-occurrence neurons are activated. When the prediction stops, the memories of the maximal prediction activity is reset by inhibitory connections. The non-occurrence signal is used to inhibit the drive that was supposed to be satisfied.

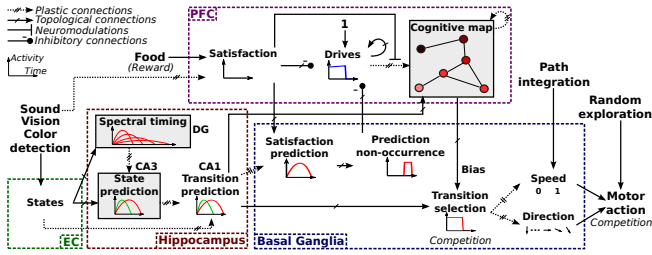


Fig. 4. Integrated model combining the timed transition learning architecture with a cognitive map navigation strategy. A detail of the new non-occurrence system represented by darker boxes is shown in fig 3.

$\alpha$  is a learning rate,  $\gamma$  a decay rate and  $\theta$  an activity threshold.

The system also detects when a predicted drive satisfaction does not occur and consequently inhibits the corresponding drive. This will modify the behavior of the robot which will attend to other objectives (i.e. seek to satisfy other active drives or explore its environment). By inhibiting the drive we tell the robot that it can currently not be satisfied. By doing so we allow the robot to resume its movement, even in the absence of any explicit stimulus terminating the waiting period. If no drive is active then there is no activity in the cognitive map and no particular transition is selected. As a result, the transition architecture does not select any particular action to perform. A small amount of noise in the action selection system will lead to the selection of a random direction. The robot will thus start a random exploration behavior. Inhibiting the different drives is a way to modulate the behavior toward different goals and to start exploring the environment if all those goals have been fulfilled or are not reachable at this time. Drives build up over time (with a speed regulated by the  $\epsilon$  parameter). This controls the frequency with which the robot needs to fulfill each objective. The complete model is shown in fig. 4.

### III. ROBOTIC EXPERIMENTS

#### A. Place preference task

Experiments are conducted on a Robosoft robulab 10 mobile robot<sup>1</sup> in an enclosed environment of 5x5m. A colored zone of 1x0.5m marks the goal location. A color detector placed under the robot allows the detection of the goal. Ultrasound sensors act as proximity detectors to provide low-level obstacle avoidance. In a preliminary phase, the robot is allowed to explore the environment for 40 minutes. A camera mounted on a pan-tilt system allows the autonomous learning and recognition of visual landmarks from the log-polar transform of 2D local views centered on the maxima of a DOG (Difference of Gaussians) filtering performed on the gradient of the image [10]. New visual place cells, represented by a constellation of visual landmarks, are autonomously learned based on a minimum activity threshold. Sensory events are characterized by changes in spatial places (based on visual information), detection of a colored zone or perception of a simulated sound representing a reward. Through random exploration, the cognitive map linking the different states is learned (fig. 5). The robot learns a topological representation of its environment and how to reach any state using the shortest path possible. The robot is controlled by the system in terms of linear and angular velocity.

The second step is to teach the robot how to perform the place preference task. The training of the robot can be facilitated by an intuitive and natural interaction with the human. In order to train the robot to wait for the sound we first need it to go to the goal location. Even though we could just wait for the robot to reach the goal by random exploration, we use a leash strapped to an artificial neck to guide the robot. This system provides an easy way for a human to interact with the robot and guide it to a location of interest. For the experiment, we set a 7s delay during which the robot has to stay in the colored zone to hear a simulated sound signaling the delivery of a food reward. The length of the delay is chosen to be significantly longer than the time it takes for the robot to cross the goal zone at its default speed. On its own, the robot cannot learn that being on the goal location could lead to receiving a reward since its default behavior (i.e. to move) does not permit it to stay long enough in the colored zone. After having guided the robot to the colored zone, we then use the low-level obstacle avoidance to make the robot stop, by placing ourselves in front of its ultra-sound sensors. Once again, this provides a natural way to interact with the robot and make it stop moving. Extra human-robot communication devices are not needed. The robot will eventually hear the sound after 7s of waiting. Proprioception about its current speed (i.e. no movement) is associated with the prediction of the sound.

In the original experiment with the rats, finding the food pellet is the actual reward and waiting at the goal location

<sup>1</sup>Neural network simulations are run on an on-board computer with an Intel Core2Duo T7500 processor running at 2.2Ghz and 2GB RAM communicating via wifi with a workstation powered by an Intel Core2Quad Q6600 running at 2.4Ghz and 4GB RAM.

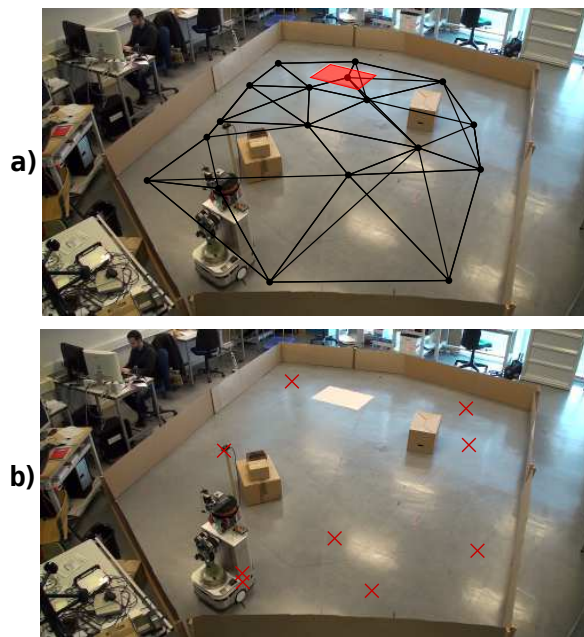


Fig. 5. **a)** Cognitive map of the environment learned during the exploration phase. The circles represent the center of the place cells and the lines represent transitions between places. **b)** Trajectories taken by the robot while planning its path to the goal using the cognitive map. Starting points are indicated by crosses and the goal by a red rectangle. The paths are not optimal (i.e. straight lines) because the algorithm selects transitions based on their distance to the goal in terms of number of nodes in the graph, which can lead to some sub-optimal decisions in the Cartesian space. Since the competition only selects one transition, the robot follows the edges of the transition graph. Smoother trajectories could be obtained by increasing the number of place cells, at the expense of memory, computational power and an increased learning time. Another solution could be to have a soft competition between transitions to select some average direction predicted by several good transitions.

is just a necessary step to obtain it. However after repeated trials, the sound of the food pellet being released after the delay must become associated with the delivery of the food. In this robotic experiment, we assume that this conditioning has already happened and that hearing the sound represents by itself a goal that needs to be satisfied. Whereas the reward should be the actual food pellet, we consider that hearing the sound is a goal in itself and has a corresponding drive that needs to be satisfied. Upon perception of the sound, the corresponding drive is inhibited and the robot starts to explore. The drive level then builds up over time until it reaches a value high enough to direct the robot toward the goal again. The time needed by the drive to reach that value simulates the time needed by the rat to find the food. The robot needs to find its way back to the goal autonomously, using the cognitive map (trajectories in fig. 5). The arrival on the goal triggers the prediction of the arrival of the sound and activates the corresponding action (stop moving). Each success reinforces this behavior.

Fig. 6 shows the speed profiles and sound prediction activity around and at the goal location for rewarded and extinction trials. It shows that the model successfully predicts the perception of the sound and controls the movement of the robot until then. In extinction trials, the prediction system is

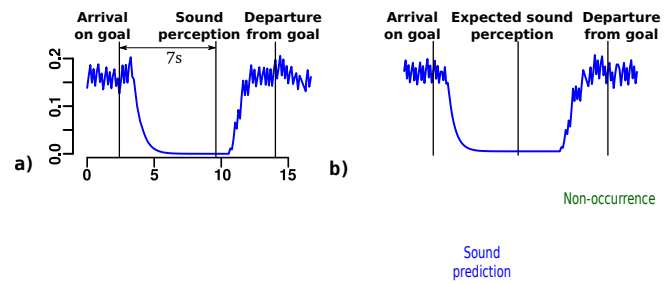


Fig. 6. Speed profile and sound prediction activity at the goal. **a)** Rewarded trials. The prediction has a bell-shaped activity with a maximum value preceding the expected timing of the satisfying stimulus. Once the sound is heard, the prediction stops and other transition cells start predicting the next stimuli. **b)** Extinction trials. The reward is not given, so no sensory stimulus signals the end of the 7s delay. Shortly after the prediction has peaked, the non-occurrence neurons are activated and start inhibiting the drive controlling the goal-oriented behavior and the robot switches to random exploration. The system has a little tolerance to delays in the occurrence of a predicted event and waits for a few seconds more after the end of the 7s waiting period before the movement of the robot is resumed. Without the non-occurrence system, the robot would wait indefinitely for the sound.

necessary for the robot to resume its movement after the end of the 7s delay. Without it the robot would wait forever for the sound, without any clue as to when it should be expecting it, and when it can consider that it will not be delivered. Additionally, the reward timing can later be modified to longer or shorter delays and the system will learn to adapt the timing of its predictions.

### B. Automated security patrol

Our model is able to perform the place preference task and gives good explanatory and predictive information about how this task is solved by rodents. However its performances are not limited to this task and they can be generalized to any navigation task involving sequences of sensory-motor behaviors with temporal constraints. One application where this type of behavior could be needed is automated security patrols. In this case the robot needs to navigate toward a sequence of checkpoints. It may have to stop at some of the checkpoints and perform some kind of behavior like looking around for intruders, for a precise amount of time, before proceeding to the next checkpoint.

Using the same neural architecture as in the place preference task, we now teach our robot to follow a given trajectory. During the first phase of the learning, the robot has absolutely no knowledge of its environment and is guided by a human using the leash. The trajectory forms a loop that the robot must complete. Place cells are autonomously learned based on a minimum activity threshold and the actions to move from one place to the next are associated to the transition between these places. When the loop is completed and the robot is back at its starting point, the leash is dropped and we let the robot reproduce the trajectory to verify that the loop was correctly learned. Because no checkpoint has been learned yet, no drive is active in the cognitive map. However the robot only knows one path and will select the

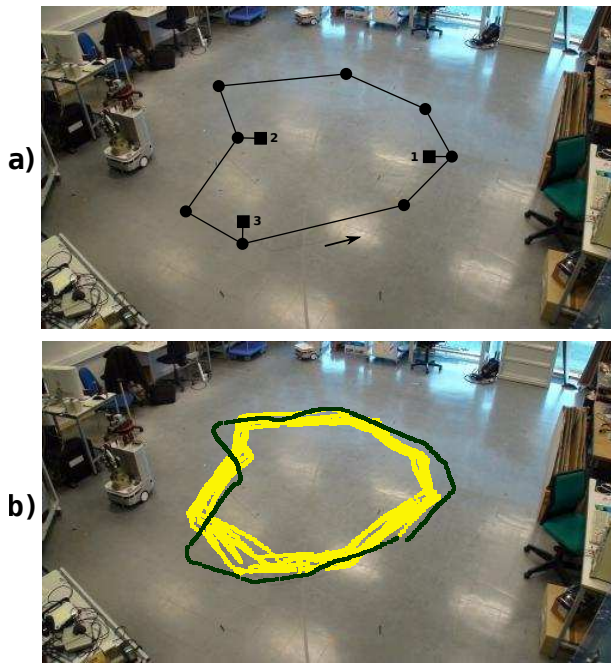


Fig. 7. **a)** Places and transitions learned for the patrol. Place field centers are signaled by black disks. Black squares represent the association between places and proprioceptive events. They point out the 3 checkpoints. **b)** Trajectories of the robot during the task. The dark trajectory corresponds to the learned path, when guided by the human. The light trajectories are the autonomous reproduction. Due to the size of place fields, the reproduced loop can be shifted compared to the original learned trajectory. This is a consequence of the prediction system that begins to predict the movement to reach the next place once it enters a new place. Since place cells can have broad place fields, this prediction can occur before the robot reaches the center of the field, which is where the robot was originally guided. The architecture offers a trade-off between the computational power required and the precision of the reproduction in the number of place cells learned on the trajectory. Indeed only movements between place cells are learned, so too few place cells will lead to a very coarse reproduction of the learned trajectory, whereas a fine reproduction can be achieved with many place cells but will lead to a longer learning phase and a memory and computationally expensive system.

only available transition in each place (the noise on the action selection system has been set to a minimal level for this task).

During the second reproduction of the patrol, the human places himself in front of the robot to make it stop. The operation is repeated at 3 different locations. The robot associates these checkpoints with the nearest place cell. The spatial accuracy of the checkpoints depends on the distance between learned place cells. The human then removes himself from the front of the robot and lets it resume the patrol. Since no visual or auditory stimulus marks the end of the waiting delay, our system needs additional information to characterize its state during and after the waiting period. We add proprioceptive information coming from odometry to the multi-modal state integration. A sensori event is triggered each time the robot resumes its movement. So in addition to transitions predicting the next place from the current checkpoint place, the robot can predict a perceptible event signaling a change from immobile to mobile. Since the robot has been immobile for the duration of the transition, the action associated is to wait without moving. The dura-

tion of the pause is learned by the system as a temporal prediction. Each checkpoint is associated with a particular drive (representing the need of the robot to inspect each checkpoint) that increases over time. When the robot arrives on a checkpoint place, if the corresponding drive is strong enough, the cognitive map will bias the selection of the waiting transition rather than the transition to the next place and the motion will stop. Once the prediction has peaked, the drive is inhibited by the non-occurrence system and the transition to the next place will win the competition and trigger its associated action: to move to the next place. The robot thus proceeds to the next phases of the patrol.

Fig. 7 shows the learned cognitive map and the trajectories for the patrol during learning and reproduction phases. In addition to the results in the previous experiment showing that a robot can learn to efficiently reach a goal by planning its path, we see here that the cognitive map can be used with good performances for the fast learning and reproduction of a trajectory. Here the learned cognitive map is rather incomplete, the robot only knows the path it has been taught. If the robot is kidnapped, by taking it and placing it far from the patrol, it will be lost. It can however be taught with the leash how to return to the patrol path and will remember it. Kidnapping the robot by placing it somewhere else on the patrol path will lead to the robot resuming the patrol and heading toward the nearest checkpoint with a high drive value. If we let the robot explore the whole environment and build a comprehensive map, it can use more direct paths between each of the checkpoints. No specific order for the checkpoints is defined in the system. The cognitive map strategy solely optimizes the paths of the robot to satisfy the strongest drives. It also takes into account the distance to the corresponding checkpoints. The frequency with which each checkpoint is visited could be controlled by adjusting the speed with which each drive rises after having been satisfied.

Table I shows the duration of the immobility of the robot at the three checkpoints. The learned duration corresponds to the time spent by the human in front of the robot. The reproduction values are calculated over 5 loops. The robot manages to reproduce the waiting behavior that was learned at the 3 checkpoints and shows consistency in the length of the reproduced waiting periods. The robot has correctly learned to suppress its movement upon entry on a particular place and resume it after a particular length of time that can be different from place to place. However another observation is that the period during which the robot is not moving can be different during the learning and reproduction phases. This is due to two reasons:

- (i) Our system learns to associate a place with a waiting

TABLE I

DURATION OF IMMOBILITY AT EACH OF THE CHECKPOINTS.

Checkpoint	1	2	3
Learned (s)	24.1	19.1	11.6
Reproduced (mean±sd)	22.3 ± 0.7	19.1 ± 0.3	15.3 ± 0.8

period of a particular duration. The duration learned by the system is the length of time between the entry on the place and the resumption of movement. When the training is done by the human, all the time spent in a place by the robot before the human stops it is added, for the learning, to the time spent waiting. To solve this issue, our system could include an additional state marking the beginning of the immobility. This would allow the learning of a timed transition between the entry on a place and the inhibition of the movement. The robot could then learn not only to wait at a particular place, but also during any movement between two places.

(ii) The prediction activity has a peak preceding the exact timing of the expected resumption of movement, learned during the first loop. However the non-occurrence neurons are only activated when the prediction activity falls back to lower values. This period between the peak of activity representing the learned timing and the drive inhibition signal, in addition to the time needed by the system and robot to transmit the motor command to the wheels, delays the restarting of the robot. Also, our predictions follow the Weber law [12]: their peak of activity preceding the expected timing of an event is less and less precise as the learn timing gets longer, as observed in animals. This explains why the long timings which are predicted well in advance by the system tend to lead to shorter reproductions.

The timing system learns continuously, not only during the first loops. The experimenter can always modify what the robot has learned. New checkpoints can be added by humans by placing themselves in front of the robot. The timings learned for existing checkpoints can be made longer in the same way. They can also be made shorter by using the leash to restart the movement of the robot before the end of the learned delay. The architecture will then learn to adapt the predicted timings.

#### IV. CONCLUSIONS AND FUTURE WORKS

We have presented a biologically plausible neural network capable of learning temporal associations between consecutive sensory events. The neural network approach allows the system to learn and adapt those timings based on task contingencies. The associations are linked to motor actions, meaning that the robot learns which actions will allow it to move from one state to another. At any time the system provides information about reachable states and about the expected timing of the arrival in a new state. A complementary network detects the non-occurrence of expected events and can modify the behavior of the robot accordingly. The latter ability forms a kind of novelty detection, allowing the robot to know when conditions in the tasks or environment may have changed. The integration of the temporal aspect of transitions between sensory events, which represents the novelty of this work, has been done with an architecture that has been successfully used in several navigation experiments, using a cognitive map [10] or Q-learning [13]. Several works have used biologically plausible models of place cells in robotic navigation and use reinforcement learning to adapt strategies to changing task contingencies [14], [15],

[16]. However these models fail to take into account the temporality of action-consequence pairings. In cases where delayed rewards which are not the immediate consequence of an action of the robot are delivered, a more complex model integrating time is required.

The non-occurrence signal really gives an internal negative feedback about the current actions of the robot. Here it is used to control the behavior of the robot by acting on the drive activity to modulate the tendency of the robot to try to reach particular goals. Yet an alternative could have been to use the signal to inhibit the currently selected transition, telling the cognitive map to look for another path to the goal. It could also be used to suppress the current navigation strategy altogether, telling the system to switch from a planning to a sensori-motor strategy. A modulation of the learning of actions, signaling that the current sensori-motor association should be suppressed because its outcome is not the predicted one, would also be an option. With such a variety of choices, we would need a high-level meta-controller to appropriately use this negative feedback. The bases of this meta-controller have already been implemented using frustration signals which built up over time if the robot stopped making progress [17]. However the system needed to be temporally calibrated for the task. With the temporal predictions, the need for this calibration would disappear.

#### REFERENCES

- [1] J. O'Keefe and J. Dostrovsky, "The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat." *Brain Res*, vol. 34, no. 1, pp. 171–175, Nov 1971.
- [2] A. Alvernhe, T. V. Cauter, E. Save, and B. Poucet, "Different ca1 and ca3 representations of novel routes in a shortcut situation." *J Neurosci*, vol. 28, no. 29, pp. 7324–7333, Jul 2008.
- [3] V. Hok, P.-P. Lenck-Santini, S. Roux, E. Save, R. U. Muller, and B. Poucet, "Goal-related activity in hippocampal place cells." *J Neurosci*, vol. 27, no. 3, pp. 472–482, Jan 2007.
- [4] B. G. Burton, V. Hok, E. Save, and B. Poucet, "Lesion of the ventral and intermediate hippocampus abolishes anticipatory activity in the medial prefrontal cortex of the rat." *Behav Brain Res*, vol. 199, no. 2, pp. 222–234, May 2009.
- [5] M. M. Quottrup, T. Bak, and R. I. Zamanabadi, "Multi-robot planning : a timed automata approach," in *Proc. IEEE Int. Conf. Robotics and Automation ICRA '04*, vol. 5, 2004, pp. 4417–4422.
- [6] H. Liu and W. Wan, "Adaptive replanning in hard changing environments," in *Proc. IEEE/RSJ Int Intelligent Robots and Systems (IROS) Conf.*, 2010, pp. 5912–5918.
- [7] R. Philippsen, S. Kolski, K. Maček, and B. Jensen, "Mobile robot planning in dynamic environments and on growable costmaps," in *Proc. IEEE Int. Conf. Robotics and Automation ICRA*, 2008.
- [8] J. P. Banquet, P. Gaussier, J. C. Dreher, C. Joulain, A. Revel, and W. Gunther, "Space-time, order and hierarchy in fronto-hippocampal system: A neural basis of personality," in *Cognitive Science Perspectives on Personality and Emotion*, 1997, pp. 123–189.
- [9] P. Andry, P. Gaussier, S. Moga, J.-P. Banquet, and J. Nadel, "Learning and communication in imitation: An autonomous robot perspective," *IEEE Transactions on Man, Systems and Cybernetics*, vol. 31, no. 5, pp. 431–442, 2001.
- [10] N. Cuperlier, M. Quoy, and P. Gaussier, "Neurobiologically inspired mobile robot navigation and planning." *Front Neurobotics*, vol. 1, p. 3, 2007.
- [11] J. Hirel, P. Gaussier, and M. Quoy, "Model of the hippocampal learning of spatio-temporal sequences," in *Artificial Neural Networks – ICANN 2010*, vol. 6354, 2010, pp. 345–351.
- [12] S. Grossberg and N. A. Schmajuk, "Neural dynamics of adaptive timing temporal discrimination during associative learning," *Neural Netw.*, vol. 2, no. 2, pp. 79–102, 1989.

- [13] J. Hirel, P. Gaussier, M. Quoy, and J.-P. Banquet, "Why and how hippocampal transition cells can be used in reinforcement learning," in *From Animals to Animats II*, vol. 6226, 2010, pp. 359–369.
- [14] A. Arleo and W. Gerstner, "Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity." *Biol Cybern*, vol. 83, no. 3, pp. 287–299, Sep 2000.
- [15] D. J. Foster, R. G. Morris, and P. Dayan, "A model of hippocampally dependent navigation, using the temporal difference learning rule." *Hippocampus*, vol. 10, no. 1, pp. 1–16, 2000.
- [16] L. Dollé, D. Sheynikhovich, B. Girard, R. Chavarriaga, and A. Guillot, "Path planning versus cue responding: a bio-inspired model of switching between navigation strategies." *Biol Cybern*, vol. 103, no. 4, pp. 299–317, Oct 2010.
- [17] C. Hasson and P. Gaussier, "Frustration as a general regulatory mechanism for motivated navigation," in *Proc. IEEE/RSJ Int Intelligent Robots and Systems (IROS) Conf*, 2010, pp. 4704–4709.