Spatial Representation and Navigation in a Bio-inspired Robot

Denis Sheynikhovich, Ricardo Chavarriaga, Thomas Strösslin,
and Wulfram Gerstner

Swiss Federal Institute of Technology,
Laboratory of Computational Neuroscience,
CH-1015 Lausanne, Switzerland
{denis.sheynikhovich, ricardo.chavarriaga,
thomas.stroesslin, wulfram.gerstner}@epfl.ch

Abstract. A biologically inspired computational model of rodent representation–based (locale) navigation is presented. The model combines visual input in the form of realistic two dimensional grey-scale images and odometer signals to drive the firing of simulated place and head direction cells via Hebbian synapses. The space representation is built incrementally and on-line without any prior information about the environment and consists of a large population of location-sensitive units (place cells) with overlapping receptive fields. Goal navigation is performed using reinforcement learning in continuous state and action spaces, where the state space is represented by population activity of the place cells. The model is able to reproduce a number of behavioral and neurophysiological data on rodents. Performance of the model was tested on both simulated and real mobile Khepera robots in a set of behavioral tasks and is comparable to the performance of animals in similar tasks.

1   Introduction

The task of self-localization and navigation to desired target locations is of crucial importance for both animals and autonomous robots. While robots often use specific sensors (e.g. distance meters or compasses), or some kind of prior information about the environment in order to develop knowledge about their location (see [1] for review), animals and humans can quickly localize themselves using incomplete information about the environment coming from their senses and without any prior knowledge. Discovery of location and direction sensitive cells in the rat's brain (see Sect. 2) gave some insight into the problem of how this self-localization process might happen in animals. It appears that using external input and self-motion information various neural structures develop activity profiles that correlate with current gaze direction and current location of the animal. Experimental evidence suggests that in many cases activity of the place and direction sensitive neurons underlies behavioral decisions, although some results are controversial (see [2], Part II for review).

The first question that we try to answer in this work is what type of sensory information processing could cause an emergence of such a location and direction sensitivity. Particular constraints on the possible mechanism that we focus on are *(i)* the absence of any prior information about the environment, *(ii)* the requirement of on-line learning from interactions with the environment and *(iii)* possibility to deploy and test the model in a real setup. We propose a neural architecture in which visual and self-motion inputs are used to achieve location and direction coding in artificial place and direction sensitive neurons. During agent-environment interactions correlations between visually– and self-motion–driven cells are discovered by means of unsupervised Hebbian learning. Such a learning process results in a robust space representation consisting of a large number of localized overlapping place fields in accordance with neuro-physiological data.

The second question is related to the use of such a representation for goal oriented behavior. A navigational task consists of finding relationships between any location in the environment and a hidden goal location identified by a reward signal received at that location in the past. These relationships can then be used to drive goal–oriented locomotor actions which represent the navigational behavior. The reinforcement learning paradigm [3] proposed a suitable framework for solving such a task. In the terms of reinforcement learning the states of the navigating system are represented by locations encoded in the population activity of the place sensitive units whereas possible actions are represented by population activity of locomotor action units. The relationships between the location and the goal are given by a state-action value function that is stored in the connections between the place and action units and learned online during a goal search phase. During a goal navigation phase at each location an action with the highest state-action value is performed resulting in movements towards the goal location. The application of the reinforcement learning paradigm is biologically justified by the existence of neurons whose activity is related to the difference between predicted and actual reward (see Sect. 2) which is at the heart of the reinforcement learning paradigm.

The text below is organized as follows. The next section describes neurophysiological and behavioral experimental data that serve as a biological motivation for our model. Section 3 reviews previous efforts in modeling spatial behavior and presents a bio-inspired model of spatial representation and navigation. Section 4 describes properties of the model and its performance in navigational tasks. A short discussion in Sect. 5 concludes the paper.

2   Biological Background

Experimental findings suggest that neural activity in several areas of the rat's brain can be related to the self-localization and navigational abilities of the animals. Cells in the hippocampus of freely moving rats termed *place cells* tend to fire only when the rat is in a particular portion of the testing environment, independently of gaze direction [4]. Different place cells are active in different parts of the environment and activity of the population of such cells encode the current

location of the rat in an allocentric frame of reference [5]. Other cells found in the hippocampal formation [6], as well as in other parts of the brain, called *head direction cells*, are active only when the rat's head is oriented towards a specific direction independently of the location (see [2], Chap. 9 for review). Different head direction cells have different preferred orientations and the population of such cells acts as an internal neural compass. Place cells and head-direction cells interact with each other and form a neural circuit for spatial representation [7].

The hippocampal formation receives inputs from many cortical associative areas and can therefore operate with highly processed information from different sensory modalities, but it appears that visual information tends to exert a dominant influence on the activity of the cells compared to other sensory inputs. For instance, rotation of salient visual stimuli in the periphery of a rat's environment causes a corresponding rotation in place [8] and head direction [6] cell representations. On the other hand, both place and head direction cells continue their location or direction specific firing even in the absence of visual landmarks (e.g. in the dark). This can be explained by taking into account integration over time of vestibular and self-movement information (that is present even in the absence of visual input), which is usually referred to as the ability to perform 'path integration'. There is an extensive experimental evidence for such 'integration' abilities of place and head direction cell populations (reviewed in [9] and [2], Chap. 9, respectively).

One of the existing hypotheses of how the place cells can be used for navigation employs a reinforcement learning paradigm in order to associate place information with goal information. In the reinforcement learning theory [3] a state space (e.g. location-specific firing) is associated with an action space (e.g. goal-oriented movements) via a state-action value function, where the value is represented by an expected future reward. This state-action value function can be learned on-line based on the information about a current location and a difference between the predicted and an actual reward. It was found that activity of dopaminergic neurons in the ventral tegmental area (VTA) of the brain (a part of the basal ganglia) is related to the errors in reward prediction [10, 11]. Furthermore these neurons project to the brain area called nucleus accumbens (NA) which has the hippocampus as the main input structure and is related to motor actions [12, 13, 14, 15]. In other words neurons in the NA receive spatial information from the hippocampus and reward prediction error information from the VTA. As mentioned before, these two types of information are the necessary prerequisites for reinforcement learning. This data supports the hypothesis that the neural substrate for goal learning could be the synapses between the hippocampus and NA. The NA further projects to the thalamus which is in turn interconnected with the primary motor cortex, thus providing a possibility that the goal information could be used to control actions. The model of navigation described in this paper is consistent with these experimental findings.

On the behavioral level, several experimental paradigms can be used to test navigational abilities of animals in the tasks in which an internal space representation is necessary for the navigation (so called *locale navigation*, see [16] for a

review of navigational strategies). Probably the most frequently used paradigm is the hidden platform water maze [17]. The experimental setup consists of a circular water pool filled with an opaque liquid and a small platform located inside the pool, but submerged below the surface of the liquid. At the beginning of each trial a rat is placed into the pool at a random location and its task is to find the platform. Since no single visual cue directly identifies the platform and the starting locations are random, animals have to remember the location of the hidden platform based on the extra-pool visual features. After several trials rats are able to swim directly to the hidden platform from any location in the pool, which indicates that they have acquired some type of spatial representation and use it to locate the platform.

Extensive lesion studies show that damage to brain areas containing place or direction sensitive cells, as well as lesions of the fornix (nerve fibers containing projections from the hippocampus to the NA) or the NA itself selectively impair navigational abilities of rats in tasks where an internal representation of space is necessary [16, 15, 18].

This and other experimental data suggest that the hippocampal formation can serve as the neural basis for spatial representation underlying navigational behavior. This hypothesized strong relation between behavior and neuro-physio-logical activity can be elaborated by means of computational models, that can in turn generate predictions testable on the level of both neurophysiology and behavior.

3   Modeling Spatial Behavior

The ability of animals to navigate in complex task-environment contexts has been the subject of a large body of research over the last decades. Because of its prominent role in memory and its spatial representation properties described above the hippocampus has been studied and modeled intensively. In the next section we review several models of the mechanisms yielding place cell activity and its role in locale navigation. In Sect. 3.2 we describe our own model in detail.

3.1   Previous Models

In this section we focus on those models which were tested in navigational tasks in real environments using mobile robots. Readers interested in theoretical and simulation models as well as in models of different types of navigation are referred to reviews of Trullier et al. [19] and Franz and Mallot [1].

Recce and Harris [20] modeled the hippocampus as an auto-associative memory which stored a scene representation consisting of the bearings and distances of the surrounding landmarks and of a goal location. The landmark bearings and distances were extracted from omnidirectional sonar scans. During a first exploration phase the scenes were stored in the memory and each stored scene was associated with a place cell. During a second goal navigation phase the currently perceived scene was compared to the scenes stored in the memory. When the scenes matched, the stored scene was activated (i.e. the place cell fired)

together with the goal location information. Once the scene was recalled, the robot moved directly to the goal. Information about the landmark positions and orientations were updated using integrated odometer signals, but the place cell activity depended only on the visual input.

Burgess et al. [21, 22] described a robotic implementation of an earlier neurophysiological model of the rat hippocampus [23]. Some place cells were shown to fire at a relatively fixed distance from the walls of a testing environment [24]. This property inspired the place recognition mechanism of the robot of Burgess et al. which visually estimated distances to the surrounding walls by detecting the position of a horizontal line at the junction of the walls and the floor in the input image. During a first exploration phase, the robot rotated on the spot at all locations of the arena to face all walls and to estimate their distances. The robot's orientation with respect to a reference direction was derived from path integration which was periodically reset by using a dedicated visual marker. A competitive learning mechanism selected a number of place cells to represent the specific wall distances for each place. In a second goal search phase, once the goal was found the robot associated four goal cells with the place cells representing four locations from which the direction towards the goal was known. During goal navigation, the goal direction could be computed from the relative activity of all goal cells using population vector technique [25].

In the model by Gaussier et al. [26, 27, 28], at each time step during exploration, a visual processing module extracted landmark information from a panoramic visual image. For each detected landmark in turn its type (e.g. vertical line in the image) and its compass bearing (their robot had a built-in magnetic compass) were merged into a single "what" (landmark type) and "where" (landmark bearing) matrix. When a place cell was recruited the active units from the "what-where" matrix were connected to it. The activity of the place cells was calculated in two steps: first, the initial activation of a place cell was determined as a product of the recognition level of a given feature and its bearing. Second, a winner-take-all mechanism reset the activities of all but the winning cell to zero. A delay in activation of the place cells between successive time steps allowed the next layer to learn place transitions: an active cell from the previous time step and an active cell from the current time step were connected to a transition cell using Hebbian learning rule. This way when a place cell was active (i.e. a place was recognized), it activated the associated transition cells thus "predicting" all possible (i.e. experienced in the past) transitions from that place. A transition selection mechanism was trained in the goal search phase: after the goal was found, the transition cells leading to the goal were activated more than others. This goal-oriented bias in the competition among possible transitions allowed the agent to find the goal.

The model of Arleo et al. [29, 30, 31] is an earlier version of the model presented in the next section. In this model position and direction information extracted from the visual input were combined with information extracted from the self-motion signals and merged into a single space representation which was then used for goal navigation. The visual processing pathway transformed a

two-dimensional camera image into a filter-based representation by sampling it with a set of orientation-sensitive filters. At each time step during an exploration phase, the agent took four snapshots, one in each cardinal direction. For each orientation, the filter activities were stored in a so called view cell. A downstream population of visual place cells combined the information from all simultaneously active view cells using a Hebbian learning rule. In the parallel self-motion processing pathway an estimation of position was performed by integrating signals from odometers. The self-motion position estimation was calibrated using the visual position estimation. Similarly, the direction estimation was performed by integrating rotations, but calibrated using a dedicated landmark (a lamp). The self-motion and visual estimations of position were then combined in "hippocampal" place cells population using Hebbian learning. The authors proposed a locale navigation system using reinforcement learning where the population of the hippocampal place cells served as a state space. Each place cell projected to four action cells, that coded for a movement in directions north, south, east and west respectively. The projection weights stored an approximated state-action value function and were modified using a reward-based learning method during a goal search phase. During navigation the action cells population vector encoded the direction of movement to the goal from any location in the environment.

3.2   A Model of Space Representation and Navigation

The computational model of the rat spatial behavior presented in this paper is an extension of the previous model by Arleo et al. (Sect. 3.1) and is able to learn a representation of the environment by exploration. Starting with no prior knowledge, the system grows incrementally based on agent–environment interaction. Information about locations visited for the first time is stored in a population of place cells. This information is subsequently used for self-localization and navigation to desired targets.

In the neural model of place cells, *allothetic* (visual) information is correlated with *idiothetic* information (rotation and displacement signals from the robot's odometers) using Hebbian learning. This yields a stable space representation where ambiguities in the visual input are resolved by the use of the idiothetic information, and a cumulative error of path integration is accounted for by using unbiased visual input.

Figure 1 presents a functional diagram of the model. Visual stimuli are encoded in the population of View Cells (VCs), which project to the population of Allothetic Place Cells (APCs) where a vision-based position estimation is acquired and to the population of Allothetic Heading Cells (AHCs) where current gaze direction is estimated from the visual input. The transformation of the visual input to the vision-based representation is a part of the allothetic pathway leading to the population of Hippocampal Place Cells (HPCs). In the second, idiothetic, pathway, displacement and rotation signals from the odometers are integrated over time to build an internal estimation of position in the Path Integration Cells (PIC) and gaze direction in the Heading Integration Cells (HIC). The path and heading integration systems allow the rat to navigate in darkness
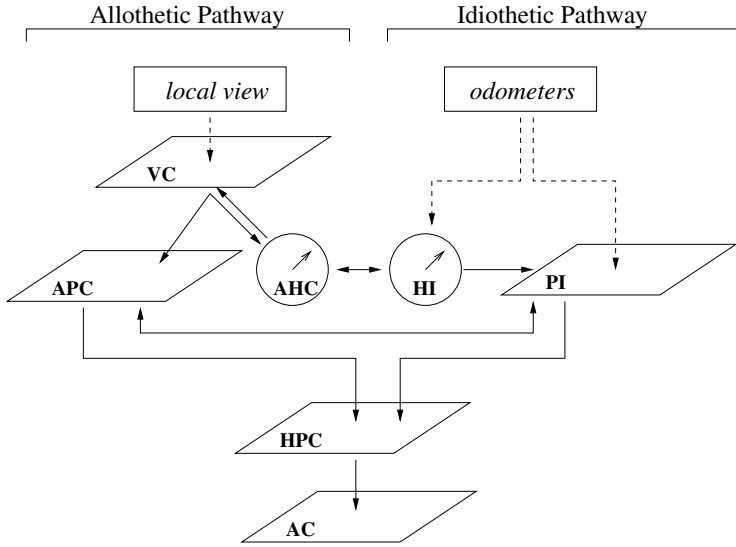
Fig. 1. Functional diagram of the model. Dashed lines denote neural transformation of a sensory input, solid lines denote projections between populations. See explanations in the text

or in the absence of visual cues. Both allothetic (APC) and idiothetic (PIC) populations project onto the HPC population where the final space representation is constructed in the form of location sensitive cells with overlapping receptive fields. Once the space representation has been learned, it can be used for navigational tasks. A direction of movement to the goal from any location in the environment is learned in the Action Cells population using temporal-difference learning technique.

The model is tested in navigational tasks using a computer simulation as well as a real Khepera robot which we refer to as 'agent' in the text below. We now discuss our model in detail.

Idiothetic Input. The idiothetic input in the model consists of rotation and displacement signals from the agent's odometers. In order to track current gaze direction we employ a population of 360 Heading Integration Cells (HIC), where each cell is assigned a preferred heading $\psi_i \in [0°, 359°]$. If $\hat{\phi}$ is the estimate of a current gaze direction, the activity of cell i from the HIC population is given by

$$r_{\psi_i}^{\mathrm{HIC}} = \exp(-(\psi_i - \hat{\phi})^2 / 2\sigma_{\mathrm{HIC}}^2) \ , \tag{1}$$

enforcing a Gaussian activity profile around $\hat{\phi}$, where σ_{HIC} defines the width of the profile. A more biologically plausible implementation of the neural network with similar properties can be realized by introducing lateral connections between the cells where each cell is positively connected to the cells with similar preferred directions and negatively connected the other cells. The attractor dy-

namics of such an implementation accounts for several properties of real head direction cells [32]. Here we employ the simpler algorithmic approach (1) that preserves network properties relevant for our model. When the agent enters a new environment arbitrary direction ψ_0 is taken as a reference direction. Whenever the agent performs a rotation, the rotational signal from the odometers is used to shift the activity blob of the HIC population. Here again a simple algorithmic approach is used where the new direction is explicitly calculated by integrating wheel rotations and a Gaussian profile is enforced around it, although more biologically plausible solutions exist [33, 34].

Having a current gaze direction encoded by the HIC population, standard trigonometric formulas can be used to calculate a new position with respect to the old position in an external Cartesian coordinate frame whenever the agent performs a linear displacement. We define Path Integration Cells (PIC) population as a two–dimensional grid of cells with predefined metric relationships, each having its preferred position $\boldsymbol{p}_i = (x_i, y_i)$ and activity

$$r_i^{\text{PIC}} = \exp(-(\boldsymbol{p}_i - \hat{\boldsymbol{p}})^2/2\sigma_{\text{PIC}}^2) \ , \tag{2}$$

where $\hat{\boldsymbol{p}} = (\hat{x}, \hat{y})$ is the estimate of position based on idiothetic information only. Origin $\boldsymbol{p}_0 = (0,0)$ is set at the entry point whenever the agent enters a new environment. The PIC population exhibits a two-dimensional Gaussian profile with width σ_{PIC} around the current position estimation.

While the agent moves through an environment the activities of HICs (1) and PICs (2) encode estimates of its position and heading with respect to the origin and the reference direction based only on the idiothetic input. They enable the agent to navigate in darkness or return to the nest location in the absence of visual cues, properties that are well known in animals [35]. The estimation of direction and position will drift over time due to accumulating errors in the odometers. Another problem is that the abstract Cartesian frame is mapped onto the physical space in a way that depends on the entry point. Both problems are addressed by combining the idiothetic input with a visual (allothetic) input and merging the two information streams into a single allocentric map.

Allothetic Input. The task of the allothetic pathway is to extract position and heading information from the external (visual) input. Based on the visually driven representation the agent should be able to recognize previously visited locations from a current local view[1]. Such a localization property can be implemented by comparing a current local view to all previously seen local views and using some similarity measure to recognize visited places (with a natural assumption that similar local views signal for spatially close locations). This comparison of the local views should take into account information about current heading that can be estimated from the relative angles between the current and all stored local views where the relative angles can in turn be computed from the amount of overlap between the local view representations.

---

[1] The term "local view" is used to denote information extracted from the visual input at a given time step.
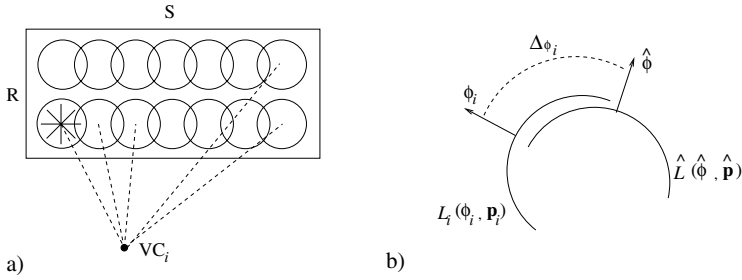
Fig. 2. Visual input and heading estimation. a: Two-dimensional panoramic image is processed by a grid of $S \times R$ points with 8 Gabor filters of different orientations at each point (filters are shown as overlapping circles, different orientations are shown only in the lower-left circle). Responses of the filters are stored in a View Cell. b: Current heading $\hat{\phi}$ can be estimated from the maximal overlap C_i between the current and a stored local views corresponding to the angular difference Δ_{ϕ_i} (3,5)

The raw visual input in the model is a two-dimensional grey-level image, received by merging several snapshots captured by the video camera of the robot into a single panoramic ($320° - 340°$) picture imitating the rat's wide view field. Note that the individual directions and the number of the snapshots are not important as long as the final image is of the required angular width (the model would perfectly suit for a single panoramic image as an input). In order to neurally represent visual input the image is sampled with a uniform rectangular grid of $S \times R$ points (see Fig. 2(a), for the results presented here we used $S = 96$ columns and $R = 12$ rows). At each point of the grid we place a set of 8 two-dimensional Gabor filters with 8 different orientations and a spatial wavelength matched to the resolution of the sampling grid. Gabor filters are sensitive to edge-like structures in the image and have been largely used to model orientation-sensitive simple cells in the visual cortex [36]. Responses F_k of $K = S \times R \times 8 = 9216$ visual filters constitute the local view information $L(\phi, \boldsymbol{p}) = \{F_k(\phi, \boldsymbol{p})\}_1^K$ that depends on the heading direction ϕ and the position $\boldsymbol{p}$ where the local view was taken.

Local views perceived by the agent are stored in the population of View Cells (VCs). At each time step a new VC i is recruited that stores the amplitudes of all current filter responses $L(\phi_i, \boldsymbol{p}_i)$. As the agent explores an environment the population of View Cells grows incrementally memorizing all local views seen so far. The place is considered to be well memorized if there is sufficient number of highly active View Cells. The information stored by the VC population can be used at each time step to estimate current heading and position as follows.

Allothetic Heading Estimation. In order to estimate current heading based on the visual input we employ a population of 360 Allothetic Heading Cells (AHC) with preferred directions uniformly distributed in $[0°, 359°]$. Suppose that a local view $L_i(\phi_i, \boldsymbol{p}_i)$ taken at position $\boldsymbol{p}_i$ in direction ϕ_i is stored by a View Cell i at time step t and the agent perceives a new local view $\hat{L}(\hat{\phi}, \hat{\boldsymbol{p}})$ at a later time step t', where $\hat{\phi}$ and $\hat{\boldsymbol{p}}$ are unknown (the local view information is time independent).

This is schematically shown in Fig. 2(b), where the arcs illustrate panoramic images (that elicit filter responses constituting the local views) with arrows showing corresponding gaze directions. In order to estimate the current heading $\hat{\phi}$ based on the stored local view L_i we first calculate the angular difference Δ_{ϕ_i} (measured in columns of the filter grid) between $\hat{L}$ and L_i that maximizes the sum of products C_i of corresponding filter values (i.e. gives maximum of the correlation function)

$$\Delta_{\phi_i} = \max_{\Delta_\phi} C_i(\Delta_\phi) \ , \tag{3}$$

$$C_i(\Delta_\phi) = \sum_s \boldsymbol{f}_i(s) \cdot \hat{\boldsymbol{f}}(s + \Delta_\phi) \ . \tag{4}$$

Here $\boldsymbol{f}_i(s)$ and $\hat{\boldsymbol{f}}(s)$ are the sets of all filter responses in vertical column s of the stored L_i and current $\hat{L}$ local views respectively, s runs over columns of the filter grid $s \in [0, S-1]$.

The estimation of the current heading $\hat{\phi}$ performed using information stored by a *single* View Cell i is now given by

$$\hat{\phi} = \phi_i + \delta_{\phi_i} \ , \quad \text{where} \tag{5}$$

$$\delta_{\phi_i} = \Delta_{\phi_i} \cdot V/S \tag{6}$$

is the angular difference measured in degrees corresponding to the angular difference measured in filter columns Δ_{ϕ_i}, V is the agent's view field in degrees.

Let us transform the algorithmic procedure (3)-(5) into a neuronal implementation. $C_i(\Delta_\phi)$ is calculated as a sum of scalar products and can hence be regarded as the output of a linear neuron with synaptic weights given by the elements of $\boldsymbol{f_i}$ applied to a shifted version of the input $\hat{\boldsymbol{f}}$. We now assume that this neuron is connected to an allothetic head direction cell with preferred direction $\psi_j = \phi_i + \delta_\phi$ and the firing rate

$$r_{\psi_j}^{\text{AHC}} = C_i(\Delta_\phi) \ , \tag{7}$$

taking into account (6). The maximally active AHC would then code for the estimation $\hat{\phi}$ of the current heading based on the information stored in VC i.

Since we have a population of View Cells (incrementally growing as the environment exploration proceeds) we can combine the estimates of *all* View Cells in order to get a more reliable estimate. Taking into account the whole VC population the activity of a single AHC will be

$$r_{\psi_j}^{\text{AHC}} = \sum_{i \in VC} C_i(\Delta_\phi) \ , \tag{8}$$

where for each AHC we sum correlations $C_i(\Delta_\phi)$ with Δ_ϕ chosen such that $\phi_i + \delta_\phi = \psi_j$. The activations (8) result in the activity profile in the AHC population. The decoding of the estimated value is done by taking a preferred direction of the maximally active cell.

Allothetic Position Estimation. As mentioned before, the idea behind the allothetic position estimation is that similar local views should signal for spatially close locations. A natural way to compare the local views is to calculate their difference

$$\Delta\tilde{L}(\phi_i, \boldsymbol{p}_i, \hat{\phi}, \hat{\boldsymbol{p}}) = \left|L_i(\phi_i, \boldsymbol{p}_i) - \hat{L}(\hat{\phi}, \hat{\boldsymbol{p}})\right| = \sum_s \left|\boldsymbol{f}_i(s) - \hat{\boldsymbol{f}}(s)\right|_1 , \qquad (9)$$

where $\boldsymbol{f}_i(s)$ and $\hat{\boldsymbol{f}}(s)$ are defined as in (4).

While exploring an environment the agent makes random movements and turns in the azimuthal plane, hence stored local views correspond to different allocentric directions. For the difference (9) to be small for spatially close locations the local views must be aligned before measuring the difference. It means that (9) should be changed to take into account the angular difference $\delta_{\phi_i} = \hat{\phi} - \phi_i$ where $\hat{\phi}$ is provided by the AHC population:

$$\Delta L(\phi_i, \boldsymbol{p}_i, \hat{\phi}, \hat{\boldsymbol{p}}) = \sum_s \left|\boldsymbol{f}_i(s) - \hat{\boldsymbol{f}}(s + \Delta_{\phi_i})\right|_1 . \qquad (10)$$

In (10) Δ_{ϕ_i} is the angular difference δ_{ϕ_i} measured in columns of the filter grid (i.e. a whole number closest to $\delta_{\phi_i} \cdot S/V$).

We set the activity of a VC to be a similarity measure between the local views:

$$r_i^{\text{VC}} = \exp\left(\frac{-\Delta L^2(\phi_i, \boldsymbol{p}_i, \hat{\phi}, \hat{\boldsymbol{p}})}{2\sigma_{\text{VC}}^2 N_\Omega}\right) , \qquad (11)$$

where N_Ω is the size of the overlap between the local views measured in filter columns and σ_{VC} is the sensitivity of the View Cell (the bigger σ_{VC} the larger is the receptive field of the cell). Each VC "votes" with its activity for the estimation of the current position. The activity is highest when a current local view is identical to the local view stored by the VC, meaning by our assumption that $\hat{\boldsymbol{p}} \approx \boldsymbol{p}_i$.

Each VC estimates current position based only on a single local view. In order to combine information from several local views, all simultaneously active VCs are connected to an Allothetic Place Cell (APC). Unsupervised Hebbian learning is applied to the connection weights between VC and APC populations. Specifically, connection weights from VC j to APC i are updated according to

$$\Delta w_{ij} = \eta \, r_i^{\text{APC}}(r_j^{\text{VC}} - w_{ij}) , \qquad (12)$$

where η is a learning rate. Activity of an APC i is calculated as a weighted average of the activity of its afferent signals.

$$r_i^{\text{APC}} = \frac{\sum_j r_j^{\text{VC}} w_{ij}}{\sum_j w_{ij}} . \qquad (13)$$

The APC population grows incrementally at each time step. Hebbian learning in the synapses between APCs and VCs extracts correlations between the View Cells so as to achieve a more reliable position estimate in the APC population.

Combined Place Code. The two different representations of space driven by visual and proprioceptive inputs are located in the APC and PIC populations respectively. At each time step the activity of PICs (2) encode current position estimation based on the odometer signals, whereas the activity of APCs (13) encode the position estimation based on local view information.

Since the position information from the two sources represent the same physical position we can construct a more reliable combined representation by using Hebbian learning.

At each time step a new Hippocampal Place Cell (HPC) is recruited and connected to all simultaneously active APCs and PICs. These connections are modified by Hebbian learning rule analogous to (12). The activity of an HPC cells is a weighed average of its APC and PIC inputs analogous to (13).

For visualization purposes the position represented by the ensemble of HPCs can be interpreted by population vector decoding [37]:

$$\hat{\boldsymbol{p}}^{\mathrm{HPC}} = \frac{\sum_j r_j^{\mathrm{HPC}} \boldsymbol{p}_j^{\mathrm{HPC}}}{\sum_j r_j^{\mathrm{HPC}}} \ , \tag{14}$$

where $\boldsymbol{p}_j^{\mathrm{HPC}}$ is the center of the place field of an HPC j.

Such a combined activity at the level of HPC population allows the system to rely on the visual information during the self-localization process at the same time resolving consistency problems inherent in a purely idiothetic system.

Goal Navigation Using Reinforcement Learning. In order to use the position estimation encoded by the HPC population for navigation, we employ a Q-learning algorithm in continuous state and action space [38, 39, 40, 3]. Values of the HPC population vector (14) represent a continuous state space. The HPC population projects to the population of N^{AC} Action Cells (AC) that code for the agent's motor commands. Each AC i represents a particular direction $\theta_i \in [0°, 359°]$ in an allocentric coordinate frame. The continuous angle θ^{AC} encoded by the AC population vector

$$\theta^{\mathrm{AC}} = \arctan\left(\frac{\sum_i r_i^{\mathrm{AC}} \cdot \sin(2\pi i / N^{\mathrm{AC}})}{\sum_i r_i^{\mathrm{AC}} \cdot \cos(2\pi i / N^{\mathrm{AC}})}\right) \tag{15}$$

determines the direction of the next movement in the allocentric frame of reference. The activity $r_i^{\mathrm{AC}} = Q(\hat{\boldsymbol{p}}^{\mathrm{HPC}}, a_i) = \sum_j w_{ij}^a r_j^{\mathrm{HPC}}$ of an Action Cell i represents a state-action value $Q(\hat{\boldsymbol{p}}^{\mathrm{HPC}}, a_i)$ of performing action a_i (i.e. movement in direction θ_i) if the current state is defined by $\hat{\boldsymbol{p}}^{\mathrm{HPC}}$. The state-action value is parameterized by the weights w_{ij}^a of the connections between HPCs and ACs.

The state-action value function in the connection values w_{ij}^a is learned according to the Q-learning algorithm using the following procedure [38]:

1. At each time step t the state-action values are computed for each action $Q(\hat{\boldsymbol{p}}^{\mathrm{HPC}}(t), a_i) = r_i^{\mathrm{AC}}(t)$.

2. Action $a = a^*$ (i.e. movement in the direction θ^{AC} defined by (15)) is chosen with probability $1-\epsilon$ (exploitation) or a random action $a = a^r$ (i.e. movement in a random direction) is chosen with probability ϵ (exploration).

3. A Gaussian profile around the chosen action a is enforced in the action cells population activity resulting in $\tilde{r}_i^{AC} = \exp(-(\theta - \theta_i)/2\sigma_{AC}^2)$, where θ and θ_i are the directions of movement coded by the actions a and a_i respectively. This step is necessary for generalization purposes and can also be performed by adding lateral connectivity between the action cells [38].

4. The eligibility trace is updated according to $e_{ij}(t) = \alpha \cdot e_{ij}(t-1) + \tilde{r}_i^{AC}(t) \cdot r_j^{PC}(t)$ with $\alpha \in [0,1]$ being the decay rate of the eligibility trace.

5. Action a is executed (along with time step update $t = t + 1$).

6. Reward prediction error is calculated as
$\delta(t) = R(t) + \gamma \cdot Q(\hat{\boldsymbol{p}}^{HPC}(t), a^*(t)) - Q(\hat{\boldsymbol{p}}^{HPC}(t-1), a(t-1))$,
where $R(t)$ is a reward received at step t.

7. Connection weights between HPC and AC populations are updated according to $\Delta w_{ij}^a(t) = \eta \cdot \delta(t) \cdot e_{ij}(t-1)$ with $\eta \in [0,1]$ being the learning rate.

Such an algorithm enables fast learning of the optimal movements from any state, in other words given the location encoded by the HPC population it learns the direction of movement towards the goal from that location. The generalization ability of the algorithm permits calculation of the optimal movement from a location even if that location was not visited during learning. Due to the usage of population vectors the system has continuous state and action spaces allowing the model to use continua of possible locations and movement directions using a finite number of place or action cells.

4 Experimental Results

In this section we are interested in the abilities of the model to *(i)* build a representation of a novel environment and *(ii)* use the representation to learn and subsequently find a goal location. The rationale behind this distinction relates to the so called latent learning (i.e. ability of animals to establish a spatial representation even in the absence of explicit rewards [41]). It is shown that having a target–independent space representation (like the HPC place fields) enables the agent to learn target–oriented navigation very quickly.

For the experiments discussed in the next sections we used a simulated as well as a real Kephera robots. In the simulated version the odometer signals and visual input are generated by a computer. The simulated odometers error is taken to be 10% of the distance moved (or angle rotated) at each time step. Simulated visual input is generated by a panoramic camera placed into a virtual environment.

4.1   Development and Accuracy of the Place Field Representation

To test the ability of the model to build a representation of space we place the robot in a novel environment (square box 100 cm.×100 cm.) and let it move in random directions incrementally building a spatial map.

Figure 3(a) shows an example of the robot's trajectory at the beginning of the exploration (after 44 time steps). During this period 44 HPCs were recruited as shown in Fig. 3(b). The cells are shown in a topological arrangement for visualization purposes only (the cells that code for close positions are not necessarily neighbors in their physical storage). After the environment is sufficiently explored (e.g. as in Fig. 3(d) after 1000 time steps), the HPC population encodes estimation of a real robot's position (Fig. 3(c)).

(a)

(b)

(c)

(d)
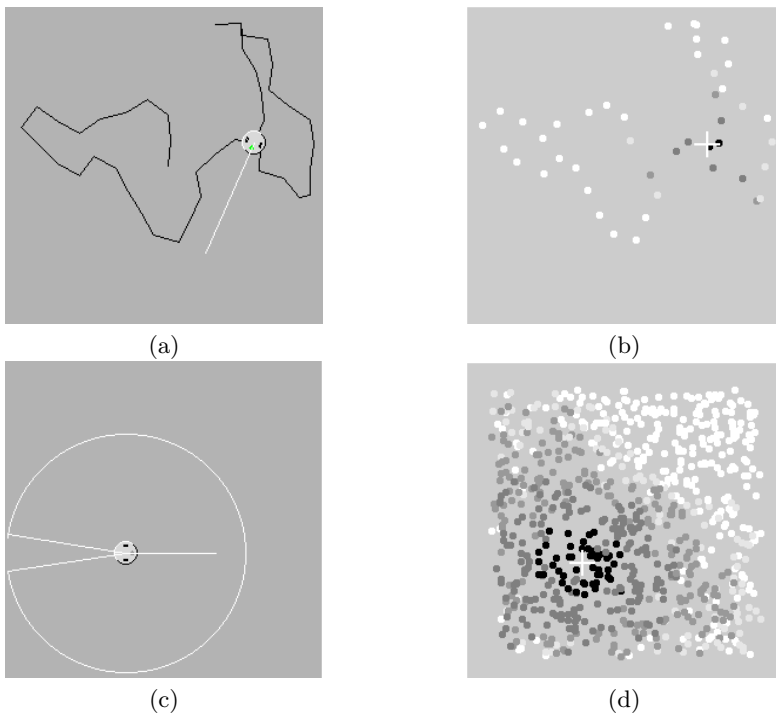
Fig. 3. Exploration of the environment and development of place cells. The grey square is the test environment. a: Exploratory trajectory of the robot after 44 time steps. Light grey circle with three dots is the robot, black line is its trajectory. The white line shows its gaze direction. b: HPCs recruited during 44 steps of exploration shown in (a). Small circles are the place cells (the darker the cell the higher its activity). c: The robot is located in the SW quadrant of the square arena heading west, white arc shows its view field (340°). d: Population activity of the HPC population after exploration while the robot's real location is shown in (c). The white cross in (b) and (d) denotes the position of the HPC population vector (14)
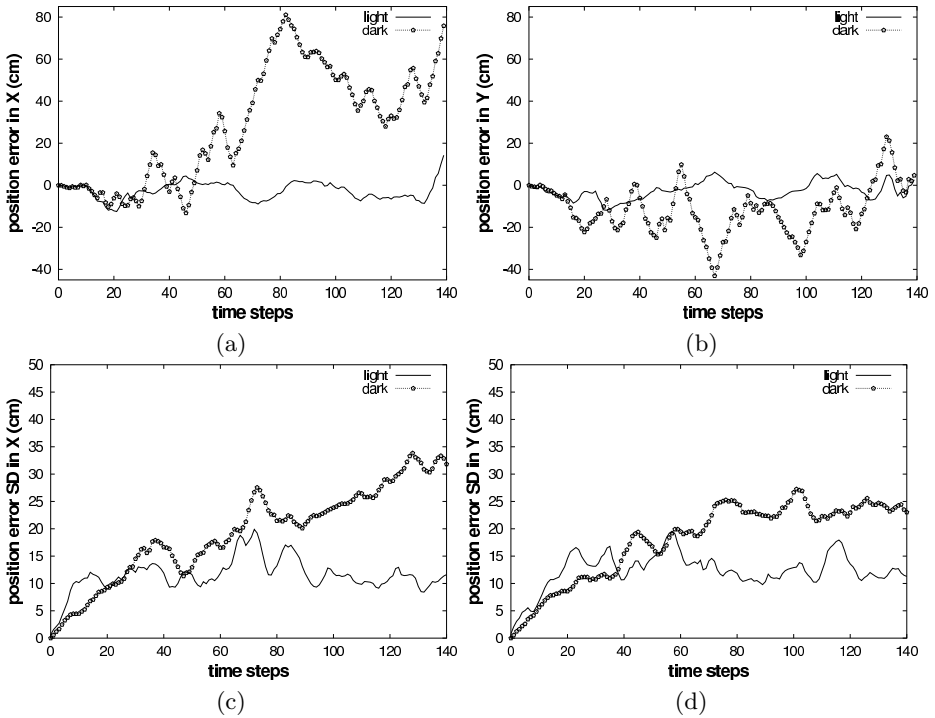
Fig. 4. a,b: Position estimation error in a single trial in X (a) and Y (b) directions with (light conditions) and without (dark conditions) taking into account the visual input. c,d: Position estimation error SD over 50 trials in X (c) and Y (d) directions in the light and dark conditions

To investigate self-localization accuracy in a familiar environment we let the robot run for 140 steps in the previously explored environment and note the error of position estimation (i.e. difference between the real position and a value of the HPC population vector (14)) at each time step in the directions defined by the walls of the box.

Figures 4(a),(b) show the error in vertical (Y) and horizontal (X) directions versus time steps ('light' conditions, solid line) in a single trial. For comparison we also plot the position estimation error in the same trials computed only by integrating the idiothetic input, i.e. without taking into account visual input ('dark' conditions, line with circles). A purely idiothetic estimate is affected by a cumulative drift over time. Taking into account visual information keeps the position error bounded.

Figure 4(c),(d) show the standard deviation (SD) of the position estimation error in light and dark conditions over 50 trials in X and Y directions (the mean error over 50 trials is approximately zero for both conditions). The error SD in light conditions is about 12 cm (that corresponds to 12% of the length of the wall).

(a)   (b)   (c)
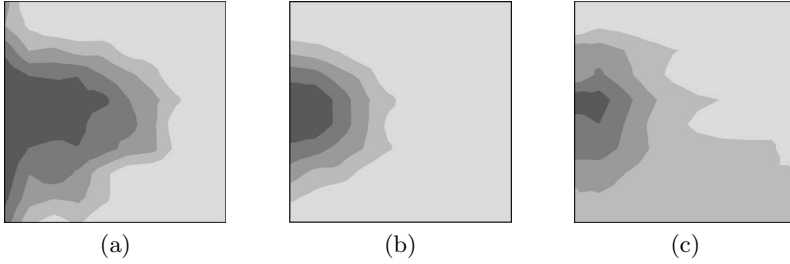
Fig. 5. a. Receptive field of a typical APC. b. Receptive field of a typical HPC. c. Receptive field of the same cell as in (b) but in dark conditions

In order to inspect the receptive fields of the place cells we let the robot systematically visit 100 locations distributed uniformly over the box area and noted the activity of a cell at each step. Contour graphs in Fig. 5(a),(b) show the activity maps for an APC and a HPC respectively. APCs tend to have large receptive fields, whereas HPC receptive fields are more compact. Each HPC combines simultaneously active PICs and APCs (see Sect. 1) allowing it to code for place even in the absence of visual stimulation, e.g. in the dark (Fig. 5(c)). This is consistent with experimental data where they found that the place fields of hippocampal place are still present in the absence of visual input [42].

4.2   Goal Directed Navigation

A standard experimental paradigm for navigational tasks that require internal representation of space is the hidden platform water maze [17]. In this task the robot has to learn how to reach a hidden goal location from any position in the environment.

The task consists of several trials. In the beginning of each trial the robot is placed at a random location in the test environment (already familiar to the robot) and is allowed to find a goal location. The position of the robot at each time step is encoded by the HPC population. During movements the connection weights between HPC and AC populations are changed according to the algorithm outlined in Sect. 1. The robot is positively rewarded each time it reaches the goal and negatively rewarded for a wall hit. The measure of performance in each trial is the number of time steps required to reach the goal (that corresponds to the amount of time required for a rat to reach the hidden platform).

After a number of trials the AC population vector (15) encodes learned direction of movement to the goal from any location $\hat{\boldsymbol{p}}^{\text{HPC}}$. A navigation map after 20 trials is shown in Fig. 6(a). The vector field representation of Fig. 6(a) was obtained by rastering uniformly over the whole environment: the ensemble responses of the action cells were recorded at 100 locations distributed over 10×10 grid of points. At each point (black dots in Fig. 6(a)) the population vector (15) was calculated and is shown as a black line where the orientation of the line corresponds to ϕ^{AC} and the length corresponds to the action value $Q(\hat{\boldsymbol{p}}^{\text{HPC}}, a^*)$.

<div style="text-align:center">(a)   (b)</div>
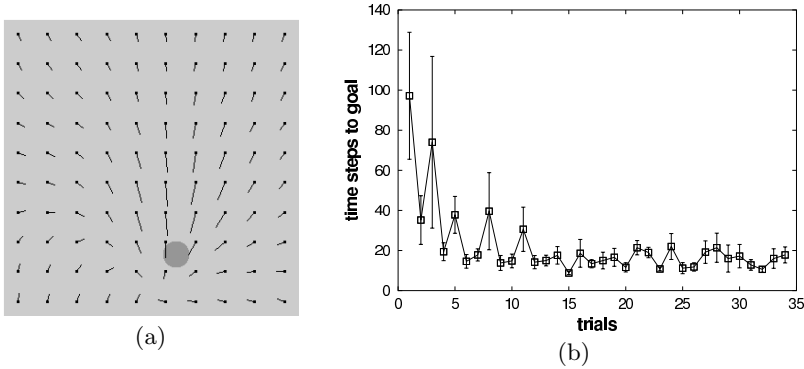
Fig. 6. a: Navigation map learned after 20 trials, dark grey circle denotes the goal location, black points denote sample locations, lines denote a learned direction of movement. b: Time to find a goal versus the number of trials

As the number of learning trials increase, the number of time steps to reach the goal decreases (Fig. 6(b)) in accordance with the experimental data with real animals [43].

5   Conclusion

The work presents a bio-inspired model of a representation–based navigation which incrementally builds a space representation from interactions with the environment and subsequently uses it to find hidden goal locations.

This model is different from the models mentioned in Sect. 3.1 in several important aspects. First, it uses realistic two–dimensional visual input which is neurally represented as a set of responses of orientation–sensitive filter distributed uniformly over the artificial retina (the visual system is similar to the one used by Arleo et al. [31], but in contrast it is not foveal in accordance with the data about the rat's visual system [44]). Second, the direction information is available in the model from the combination of visual and self-motion input, no specific compass or dedicated orientational landmark are used. Third, as in the model by Arleo et al. the integration of the idiothetic information (i.e. path integration) is an integrative part of the system that permits navigation in the dark and supports place and head direction cells firing in the absence of visual input.

The model captures some aspects of related biological systems on both behavioral (goal navigation) and neuronal (place cells) levels. In experimental neuroscience the issue of relating neuro-physiological properties of neurons to behavior is an important task. It is one of the advantages of modeling that potential connections between neuronal activity and behavior can be explored systematically. The fact that neuro-mimetic robots are simpler and more experimentally transparent than biological organisms makes them a useful tool to check new hypotheses and make predictions concerning the underlying mechanisms of spa-

tial behavior in animals. On the other hand, a bio-inspired approach in robotics may help to discover new ways of building powerful and adaptive robots.

References

1. Franz, M.O., Mallot, H.A.: Biomimetic robot navigation. Robotics and Autonomous Systems **30** (2000) 133–153
2. Jeffery, K.J., ed.: The neurobiology of spatial behavior. Oxford University Press (2003)
3. Sutton, R., Barto, A.G.: Reinforcement Learning - An Introduction. MIT Press (1998)
4. O'Keefe, J., Dostrovsky, J.: The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. Brain Research **34** (1971) 171–175
5. Wilson, M.A., McNaughton, B.L.: Dynamics of the hippocampal ensemble code for space. Science **261** (1993) 1055–1058
6. Taube, J.S., Muller, R.I., Ranck Jr., J.B.: Head direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. Journal of Neuroscience **10** (1990) 420–435
7. Knierim, J.J., Kudrimoti, H.S., McNaughton, B.L.: Place cells, head direction cells, and the learning of landmark stability. Journal of Neuroscience **15** (1995) 1648–1659
8. Muller, R.U., Kubie, J.L.: The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. Journal of Neuroscience **7** (1987) 1951–1968
9. McNaughton, B.L., Barnes, C.A., Gerrard, J.L., Gothard, K., Jung, M.W., Knierim, J.J., Kudrimoti, H., Qin, Y., Skaggs, W.E., Suster, M., Weaver, K.L.: Deciphering the hippocampal polyglot: the hippocampus as a path integration system. J Exp Biol **199** (1996) 173–85
10. Schultz, W., Dayan, P., Montague, P.R.: A neural substrate of prediction and reward. Science **275** (1997) 1593–1599
11. Schultz, W.: Predictive Reward Signal of Dopamine Neurons. Journal of Neurophysiology **80** (1998) 1–27
12. Freund, T.F., Powell, J.F., Smith, A.D.: Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. Neuroscience **13** (1984) 1189–215
13. Sesack, S.R., Pickel, V.M.: In the rat medial nucleus accumbens, hippocampal and catecholaminergic terminals converge on spiny neurons and are in apposition to each other. Brain Res **527** (1990) 266–79
14. Eichenbaum, H., Stewart, C., Morris, R.G.M.: Hippocampal representation in place learning. Journal of Neuroscience **10(11)** (1990) 3531–3542
15. Sutherland, R.J., Rodriguez, A.J.: The role of the fornix/fimbria and some related subcortical structures in place learning and memory. Behavioral and Brain Research **32** (1990) 265–277
16. Redish, A.D.: Beyond the Cognitive Map, From Place Cells to Episodic Memory. MIT Press-Bradford Books, London (1999)
17. Morris, R.G.M.: Spatial localization does not require the presence of local cues. Learning and Motivation **12** (1981) 239–260
18. Packard, M.G., McGaugh, J.L.: Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. Behavioral Neuroscience **106(3)** (1992) 439–446

19. Trullier, O., Wiener, S.I., Berthoz, A., Meyer, J.A.: Biologically-based artificial navigation systems: Review and prospects. Progress in Neurobiology **51** (1997) 483–544
20. Recce, M., Harris, K.D.: Memory for places: A navigational model in support of Marr's theory of hippocampal function. Hippocampus **6** (1996) 85–123
21. Burgess, N., Donnett, J.G., Jeffery, K.J., O'Keefe, J.: Robotic and neuronal simulation of the hippocampus and rat navigation. Phil. Trans. R. Soc. Lond. B **352** (1997) 1535–1543
22. Burgess, N., Jackson, A., Hartley, T., O'Keefe, J.: Predictions derived from modelling the hippocampal role in navigation. Biol. Cybern. **83** (2000) 301–312
23. Burgess, N., Recce, M., O'Keefe, J.: A model of hippocampal function. Neural Networks **7** (1994) 1065–1081
24. O'Keefe, J., Burgess, N.: Geometric determinants of the place fields of hippocampal neurons. Nature **381** (1996) 425–428
25. Georgopoulos, A.P., Kettner, R.E., Schwartz, A.: Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. Neuroscience **8** (1988) 2928–2937
26. Gaussier, P., Leprêtre, S., Joulain, C., Revel, A., Quoy, M., Banquet, J.P.: Animal and robot learning: Experiments and models about visual navigation. In: 7th European Workshop on Learning Robots, Edinburgh, UK (1998)
27. Gaussier, P., Joulain, C., Banquet, J.P., Leprêtre, S., Revel, A.: The visual homing problem: An example of robotics/biology cross fertilization. Robotics and Autonomous Systems **30** (2000) 155–180
28. Gaussier, P., Revel, A., Banquet, J.P., Babeau, V.: From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. Biol Cybern **86** (2002) 15–28
29. Arleo, A., Gerstner, W.: Spatial cognition and neuro-mimetic navigation: A model of hippocampal place cell activity. Biological Cybernetics, Special Issue on Navigation in Biological and Artificial Systems **83** (2000) 287–299
30. Arleo, A., Smeraldi, F., Hug, S., Gerstner, W.: Place cells and spatial navigation based on 2d visual feature extraction, path integration, and reinforcement learning. In Leen, T.K., Dietterich, T.G., Tresp, V., eds.: Advances in Neural Information Processing Systems 13, MIT Press (2001) 89–95
31. Arleo, A., Smeraldi, F., Gerstner, W.: Cognitive navigation based on nonuniform gabor space sampling, unsupervised growing networks, and reinforcement learning. IEEE Transactions on Neural Networks **15** (2004) 639–652
32. Zhang, K.: Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. Journal of Neuroscience **16(6)** (1996) 2112–2126
33. Arleo, A., Gerstner, W.: Spatial orientation in navigating agents: Modeling head-direction cells. Neurocomputing **38–40** (2001) 1059–1065
34. Skaggs, W.E., Knierim, J.J., Kudrimoti, H.S., McNaughton, B.L.: A model of the neural basis of the rat's sense of direction. In Tesauro, G., Touretzky, D.S., Leen, T.K., eds.: Advances in Neural Information Processing Systems 7, Cambridge, MA, MIT Press (1995) 173–180
35. Etienne, A.S., Jeffery, K.J.: Path integration in mammals. HIPPOCAMPUS (2004)
36. Daugman, J.G.: Two-dimensional spectral analysis of cortical receptive field profiles. Vision Research **20** (1980) 847–856
37. Pouget, A., Dayan, P., Zemel, R.S.: Inference and computation with population codes. Annu. Rev. Neurosci. **26** (2003) 381–410

38. Strösslin, T., Gerstner, W.: Reinforcement learning in continuous state and action space. In: Artificial Neural Networks - ICANN 2003. (2003)
39. Foster, D.J., Morris, R.G.M., Dayan, P.: A model of hippocampally dependent navigation, using the temporal difference learning rule. Hippocampus **10(1)** (2000) 1–16
40. Doya, K.: Reinforcement learning in continuous time and space. Neural Computation **12** (2000) 219–245
41. Tolman, E.C.: Cognitive maps in rats and men. Psychological Review **55** (1948) 189–208
42. Quirk, G.J., Muller, R.U., Kubie, J.L.: The firing of hippocampal place cells in the dark depends on the rat's recent experience. Journal of Neuroscience **10** (1990) 2008–2017
43. Morris, R.G.M., Garrud, P., Rawlins, J.N.P., O'Keefe, J.: Place navigation impaired in rats with hippocampal lesions. Nature **297** (1982) 681–683
44. Hughes, A.: The topography of vision in mammals of contrasting life style: Comparative optics and retinal organisation. In Crescitelli, F., ed.: The Visual System in Vertebrates. Volume 7/5 of Handbook of Sensory Physiology. Springer-Verlag, Berlin (1977) 613–756