Theory Biosci. (2012) 131:129–137
DOI 10.1007/s12064-011-0141-0

ORIGINAL PAPER

Variants of guided self-organization for robot control

Georg Martius · J. Michael Herrmann

Received: 31 May 2010 / Accepted: 15 October 2010 / Published online: 25 November 2011

Abstract Autonomous robots can generate exploratory behavior by self-organization of the sensorimotor loop. We show that the behavioral manifold that is covered in this way can be modified in a goal-dependent way without reducing the self-induced activity of the robot. We present three strategies for guided self-organization, namely by using external rewards, a problem-specific error function, or assumptions about the symmetries of the desired behavior. The strategies are analyzed for two different robots in a physically realistic simulation.

Keywords Guided self-organization · Autonomous robots · Homeokinesis · Machine learning

Introduction

Self-organization of robot control selects behavioral modes that are simultaneously optimized for sensitivity and predictability. The resulting behavior is characterized by on-going exploration or by a refinement of those behavioral traits which have come to be called 'natural' for a particular robot in a particular environment. Animals, including humans, acquire their behavioral repertoire in a similar way, elementary behaviors are brought forth autonomously and are further refined during the whole life span. Nevertheless, the effects due to learning that modulate the self-organized behavior are often not intrinsically produced. Animals can learn by imitation or by downright teaching through superior fellows. In addition, behavior is subject to the dictate of drives that have causes outside the biomechanical interplay of body and environment. Humans derive goals for their own behavior from rational reasoning. Each of these incentives for behavioral adaptation is an interesting subject for study in behavioral science, while the relation between such higher forms of learning and primordial self-organization of the behavioral elements remains elusive. In robotics the situation is only slightly different. Although there exist promising examples (Verschure et al. 1992; Kelso 1995; Herrman 2001; Der et al. 2006), self-organization of behavior is still a field of active exploration. Further questions such as the interaction of learning by self-organization and learning by supervision or by external reinforcement are just starting to gain scientific interest.

Usually, goal-directed behavior is achieved by directly optimizing the parameters of a control program such that the goal is approached more closely. The learning system must receive information about whether or not the behavior actually approaches the goal. This information may be available via a reward signal in reinforcement learning (Sutton et al. 1998) or by a fitness function in evolutionary algorithms (Nolfi et al. 2001). We will allow for different types of goal-related information when aiming at a combination of self-organizing control with external signals or drives. For this combination the term guided self-organization (GSO) was extended (Martius et al. 2007; Prokopenko 2009) beyond earlier usages in nanotechnology (Choi et al. 2005) and swarm robotics (Rodriguez 2007). In this general perspective, GSO is the combination of goal-

G. Martius (✉)
Bernstein Center for Computational Neuroscience
and Max Planck Institute for Dynamics and Self-Organization,
Bunsenstr. 10, 37073 Göttingen, Germany
e-mail: martius@mis.mpg.de

J. M. Herrmann
Institute for Perception, Action and Behaviour, School
of Informatics, University of Edinburgh, 10 Crichton Street,
Edinburgh EH8 9AB, UK
e-mail: michael.herrmann@ed.ac.uk

oriented learning and developmental self-organization. Each of the two paradigms brings about its particular benefits and GSO aims at combining them in a useful manner. Self-organizing systems tend to have a high tolerance against failures and degrade gracefully, which is an advantage that should not be given up when developing systems for practical applications. In statistical learning an analogous approach was proposed based on known symmetries or hints (Abu-Mostaf 1995).

Although a wider context could be interesting as well, we will be dealing here with a specific approach to self-organizing control, namely homeokinetic learning (Der 2001). What can we expect from a guided homeokinetic controller? It has been shown earlier (Der and Liebscher 2002; Der et al. 2002, 2006) that a variety of behaviors can emerged from the principle of homeokinesis. The process of self-organization has quickly structured the space of action sequences into a set of behaviors that show a coherent sensorimotor dynamics of the particular robot in its environment. The goal is now to shape the self-organization process to produce specific behaviors within a short time. Part of the idea is to channel the exploration of the homeokinetic controller around certain desired behaviors, such that modes can be found which fit even better to the particular robotic device. This is especially important in high-dimensional systems where the self-organized search for behaviors can take a long time and it is not guaranteed that all possible behaviors are visited in finite time. With additional soft constraints we can expect to achieve potentially useful behaviors even in high-dimensional robotic systems.

What are specifically the challenges of guided self-organization? Prokopenko (2008) has summarized the differences between self-organization and the realization of a particular function in engineering as follows:

> In fact, one may argue that the notions of design and self-organization are contradictory: the former approach often assumes a methodical step-by-step planning process with predictable outcomes, whereas the latter involves non-deterministic spontaneous dynamics with emergent features.

The challenge is to combine both in a favorable way yielding a system that self-organizes such that the desired function emerges and the properties like reorganization and graceful handling of failures of the self-organizing system remain.

In this article, we will discuss three mechanisms of guidance. The first one uses online reward signals to shape the emerging behaviors and is briefly discussed in Guided self-organizing control, section. A second mechanism for guiding allows for the incorporation of supervised learning signals, e.g., specific nominal motor commands, which we call teaching signals. Using distal learning (Jordan 1992) we study the use of teaching signals in terms of sensor values. This approach and a third mechanism that allows for the specification of cross-motor teaching are presented here for the first time and are given most of the available space. In particular the latter will be proved an effective and simple way to introduce useful constraints into the system and facilitate the unsupervised development of specific behaviors.

Self-organized closed-loop control

Self-organizing control for autonomous robots can be achieved by establishing an intrinsic drive toward behavioral activity (Der 2001). We will formulate an appropriate control law within a dynamical systems approach, i.e., we consider the sensor values, motor actions and possibly internal parameters of a robot as the state of a dynamical system. In order to formulate an adaptation rule for the parameter-dependent controller of the robot a simplified version of this dynamical system is obtained in form of a mathematical model. In this context the controller represents a function that maps sensory inputs to motor activations. The actuators driven in this way change the relation of the robot and its environment and typically also the sensory inputs are changed. In the following time step new sensor values are measured and so forth, such that the system forms a closed loop.

The self-evaluation of the behavior of the robot as well as planning, complex control, and high-level optimization require an internal model of how changes of sensor values are caused by motor actions. For an open-ended acquisition of behaviors this model must be adaptive which is assumed here to be realized by the gradual improvement of a parameter vector, i.e., the system is fully determined for a concrete environment by the parameters of the controller of the robot, the parameters of the model, and the initial state. Both controller and internal model are implemented as artificial neural networks. The update of all parameters is achieved by minimizing certain functionals of the mismatch between the prediction and the actual sensor values, i.e., the prediction error. The internal model is adapted by a direct minimization of this error using a gradient descent algorithm.

It is, however, not advisable to train also the controller parameters by the same direct criterion, because each behavior with unchanging or well predictable sensor values becomes a stable attractor. Therefore, usually only trivial behaviors are acquired, the robot may be doing nothing for instance. Learning according to this principle therefore tends to result in an impoverishment of the behavior. In fact, the direct minimization of prediction errors in a dynamical context can be related to homeostasis (Cannon 1939), but either the drive toward general activity must arise from elsewhere (Di Paolo 2003) or the "trivial" behavior itself must be desirable such as in walking.

Interestingly, already a seemingly minor re-interpretation of the prediction error, as proposed by the homeokinetic principle (Der 2001; Der and Liebscher 2002; Der et al. 2002, 2006), solves the behavioral impoverishment problem. The new objective function is based on the postdiction error—the difference between observed and re-estimated sensor values. Once new sensor values are available previous sensor values can be re-estimated using the inversion of the mathematical model. This re-estimation is not perfect, hence we have a mismatch which is to be minimized. Note, that if these re-estimated (virtual) sensor values were actually observed instead of the true ones then optimal prediction would have been possible. In this way, the prediction error is implicitly also minimized. What seems to be a mathematical nuisance has in fact essential consequences for the stability of the learning scheme. A schematic view of the homeokinetic controller is displayed in Fig. 1.

We have seen that the minimization of the standard prediction error, the forward error, leads to stable and often trivial behaviors. The postdiction (or backward) error, which is given by the difference between the previously measured sensor values and their current re-estimates, opens a different perspective for learning. In order to obtain the backward error the sensorimotor loop function is formally inverted. Practically, a linear approximation is sufficient where the inverted Jacobian matrix transforms the prediction error into the postdiction error.

The usage of the backward error for the update of the forward dynamics implies a virtual inversion of time which is known to reverse the stability properties of the system. The effect is, however, slightly more subtle: We assume that the prediction error at one time step is continuously related to the error at the next time step. Therefore, the minimization of a the postdiction error still keeps the actual prediction error within certain bounds. When using the backward error, actually the relation between the two types of errors is affected. Owing to the adaptation a small backward error will tend to produce a comparatively large forward error which is manifested behaviorally in the sensitivity of the robot with respect to environmental stimuli. Since, however, the predictability is still optimized at the same time, the approach mediates the two seemingly contradicting goals of predictability and sensitivity.

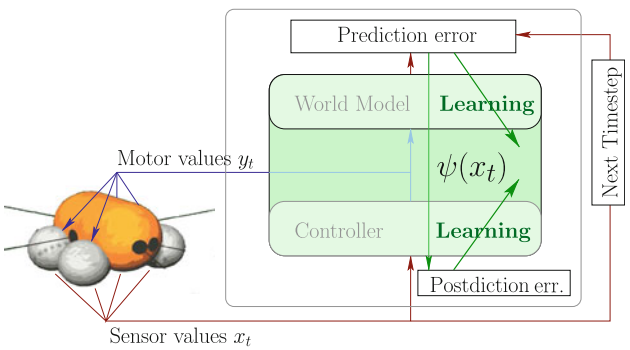

Fig. 1 Homokinetic controller in the sensorimotor loop. The postdiction error is obtained from the prediction error by inverting the sensorimotor loop ψ, cf. Box 1

Box 1

The sensorimotor loop is modeled by a function ψ which contains the controller and world model. It maps current sensor values (x_t) to the next sensor values (x_{t+1}):

$$x_{t+1} = \psi(x_t) + \xi_{t+1} \quad (1)$$

where ξ is the prediction misfit. The parameters of the world model are adapted to minimize the prediction error $E_{\text{Pred}} = \|\xi_{t+1}\|$. The controller parameters C are adapted to minimize the postdiction error

$$E_{\text{Post}} = \left\|L_t^{-1}\xi_{t+1}\right\| \quad \text{with } L_{t,ij} = \frac{\partial\psi(x_t)_i}{\partial x_{t,j}} \quad (2)$$

where L_t is the Jacobian matrix of the sensorimotor loop. In this way we find $C_{t+1} = C_t - \varepsilon_C \frac{\partial}{\partial C} E_{\text{Post}}$ where a typical value of the learning rate is $\varepsilon_C = 0.1$. This gives rise to a simultaneous dynamics of the state and of the parameters

We should note that the postdiction error can be minimized both by adapting the behavior and by improving the model. Therefore, model and controller behave complementary as the model dampens the controller while the controller activates the model by sensitizing the behavior. In the experiments this dynamic complementarity is seen to produce a rich repertoire of behaviors that explore the manifold of interactions between the robot and its environment.

The dynamics specified in Box 1 produces an itinerant trajectory in parameter-space corresponding to a sequence of behaviors of the robot. These behaviors are, however, waxing and waning and their time span and transitions are hard to predict. Although all emerging behaviors are in a sense 'natural' for the robot in the interaction with its environment, only some of the behaviors are potentially useful, interesting or beneficial. In the following we will present a mechanism that still exploits the potential richness of the behavioral manifold but biases or guides the self-organization of the robotic behaviors toward desirable behaviors.

Let us now consider an application of the homeokinetic controller. For that we use a simulated robot named the SPHERICAL which is of a relatively simple design, however, involving a complex control problem, see Fig. 2. This robot has a spherical body and is actuated by three internal weights that are movable along orthogonal axes. Thus any change in the positions of the weights results in a change of the center of mass of the robots and thus in a certain rolling movement. The control of the system has to take into account inertia effects and the non-trivial relation between motor actions and body movements.

The controller determines the new target position of the three weights along their axes and simulated motors are used to move the weights to these positions. Initially the robot is placed on even ground and does not move. As a consequence of the homeokinetic learning rule the controller becomes more and more sensitive to its sensor value changes. The first movements are due to the amplification of small noisy fluctuations until a more coherent physical movement develops. Shortly afterwards a regular rolling behavior is executed which breaks down infrequently to give way for different movement patterns. In particular the rolling modes around one of the internal axes are occurring, see Fig. 3.

To summarize, the homeokinetic controller produces body-related behaviors by self-organization of the sensorimotor dynamics of the robot including physical states and internal parameters. Given minimal sensory noise, the behaviors emerge for arbitrary initial conditions.[1] The behaviors are waxing and waning due to the ongoing re-organization process.

How can the controller achieve this apparently effective control strategy without specific information about the robot and its environment? It is important to realize that also a fixed closed-loop controller is able to create a single of the observed behaviors by a constant mapping of sensor values to motor actions. Because the self-organizing controller explores the relevant regions in the parameter space it arrives at the appropriate region and remains there longer than elsewhere. This is due to small prediction errors in this region and the high stability of the physical dynamics, while elsewhere larger errors also cause faster jumps through the parameter space. If, in particular, rolling movements are desired, the natural preferences of the learning rule are nearly optimal. It is an interesting option to use the self-referential learning dynamics of the present approach with a modification of the system in order to change the 'natural' behaviors. In this way, a much larger class of objectives can be achieved while still maintaining a self-organization of the dynamics. Illustrative examples such as curved rolling, rotation around one particular axis and other will be shown in the following.

Guided self-organizing control

How can we guide the learning dynamics such that a given goal is realized by the self-organizing process? One option was already suggested above in the experiment with the SPHERICAL robot where the variations of the error cause the system to stay in different regions of the parameter space for different durations such that certain behaviors are observed more often than others. If we include a reward signal into the learning rule of the robot we can explicitly modify these frequencies and obtain more of a desired and less of an undesired behavior. Because the prediction error acts as a factor in the learning rule, well predictable behaviors persist longer than the more chaotic ones. Therefore, weighting the error value according to the desirability of a behavior can increase the duration of rewarded behaviors while punished behaviors can be suppressed. When applying this method to the SPHERICAL robot we can e.g. achieve fast locomotion by rewarding for high velocity and obtain curved driving and spinning modes when rewarding for rotational velocity around the upwards axis, see (Martius et al. 2007) for details.

A second and more stringent form of guidance will be studied in this article. We will formulate the problem in terms of problem-specific error functions (PSEF) that indicate an external goal by minimal values. A trivial example of such an error function is the difference between externally defined and actually executed motor actions. This is a standard control problem which, however, becomes hard if the exploratory dynamics is not abandoned.

Guided self-organization (GSO) focuses on this interplay between the explorative dynamics implied by homeokinetic learning and the additional drives. The challenge in the combination of a self-organizing system with external goals becomes clear when recalling the characteristics of a self-organizing system. One important feature is the spontaneous breaking of symmetries of the system. This is a prerequisite for spontaneous pattern formation and is usually achieved by self-amplification, i.e., small noisy perturbations cause the system to choose one of several symmetric options while the intrinsic dynamics then causes the system to settle into this asymmetric state. A nonlinear stabilization of the self-amplification forms another ingredient of self-organization. These two conditions which we will call our working regime, are to be met for a successful guidance of a self-organizing system. There are a number of ways to guide the homeokinetic controller which we will discuss in the following.

Guidance by teaching

Next we will describe how problem-specific error functions (PSEFs) can be integrated into the homeokinetic approach and consider a few illustrative examples. If a PSEF depends functionally on the controller parameters it could be minimized by gradient descent in the same way as the homeokinetic error function. However, if the learning rule

[1] There are some formal requirements on the parameters, for instance that the determinant of the Jacobian matrix of the sensorimotor loop is positive.

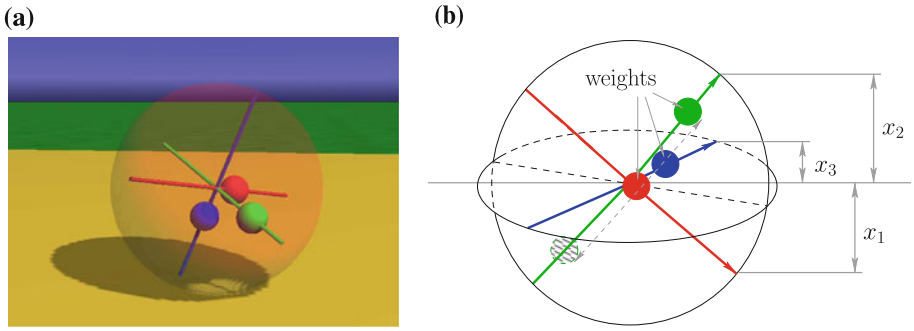

Fig. 2 SPHERICAL robot: driven by weights and equipped with axis-orientation sensors. **a** Screen shot from a physically realistic simulation. The ball-shaped weights are moved by actuator along the axes; **b** Schematic view of the robot with axis-orientation sensors (x_i)

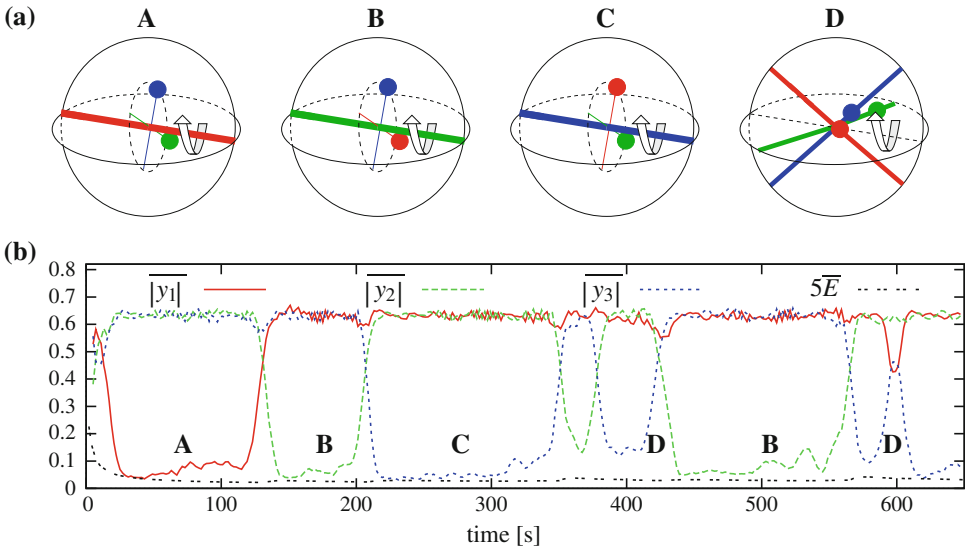

Fig. 3 SPHERICAL robot exploring its behavioral capabilities. **a** Sketch of four typical behaviors *A*–*D*, namely the rolling mode around the three internal axis (*A*–*C*) and around any another axis (*D*); **b** Amplitudes of the motor value oscillations ($y_{1\ldots3}$) and the time loop error (*E*) averaged over 10 s (scaled for visibility). Corresponding behaviors are indicated with *letters* (*A*–*D*)

is based on the sum of the homeokinetic error and the PSEF then typically either the PSEF dominates the behavior or is not effective at all. It is not obvious how to determine a fixed weighting of the two contributions that leads to a pursuit of the goal while still maintaining exploratoriness. One reason is that the nonlinearities in the postdiction error cause the gradient to vary over orders of magnitude. A solution to this problem consists in scaling the contribution of the PSEF according to the response strength of the sensorimotor loop including the nonlinearities, such that gradient of the homeokinetic error and the gradient of the PSEF have a comparable effect. This scaling can be achieved by a natural gradient (Amari 1998) in the parameter update where the Jacobian matrix of the sensorimotor loop is used as a metric, see Box 2.

To start we consider the simple problem of a wheeled robot that is to follow predefined motor actions called teaching signals which are given externally with respect to the self-organizing system. Since the controller is acting in a closed loop, we have to capture the correct input-output mapping rather than to learn a sequence actions. Therefore, we can define the PSEF as the mismatch between motor teaching signals and the actual motor values, see Box 2.

Box 2

The controller parameters *C* are updated by gradient descent with learning rate ε_C.

$$C_{t+1} = C_t - \varepsilon_C \frac{\partial E_{\text{Post}}}{\partial C} - \varepsilon_C \gamma \frac{\partial E_{\text{G}}}{\partial C} \left(L_t L_t^\top\right)^{-1} \tag{3}$$

The guidance factor $\gamma \geq 0$ is usually small since sensitivity to guidance is assured by the homeokinetic term E_{Post} defined in Eq. 2 in Box 1 on page 4. The last term contains the problem specific error function

$$E_{\text{G}} = \|y_t^{\text{G}} - y_t\|^2 \tag{4}$$

The latter have a direct functional dependency on the controller parameters, such that the gradient descent can be performed. In order to evaluate this mechanism of guidance we will analyze the TWOWHEELED robot which is a platform similar to the Khepera robot. The two wheels are driven by motors that can rotate in both directions. The motor signals determine the desired rotational velocity which is then checked by two sensors for the actual wheel velocities. The difference between the actual and desired velocity can be used by the robot to detect the presence of an obstacle. As a test of guidance we provide the controllers of the two motors with an oscillating teaching signal. To be integrated

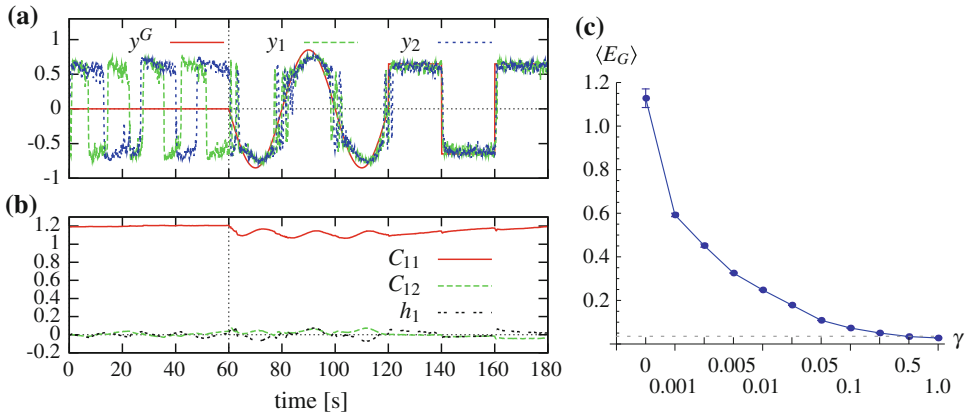

Fig. 4 TwoWheeled robot controlled with homeokinetic controller and motor teaching signals. **a** The teaching signals y^G (identical in both components) are followed partially by the motor values $y_{1,2}$ after teaching was switched on with $\gamma = 0.01$ at 60 s. **b** Time evolution of the controller parameters affecting the first motor is shown to illustrate that only little changes are necessary, however, the adaptations do not vanish; **c** Average value of the PSEF E^G (for five experiments each 5 min long) in dependence of γ (note the logarithmic scale). The noise level (*gray dotted line*) is reached at $\gamma = 1$. Parameters: (a, b) $\gamma = 0.01$, update rate 100 Hz

more easily it is at first smoothly sine shaped, while at an intermediate phase it is turned into a step function. The resulting behavior is a mixture between the taught behavior and self-organized dynamics. The teaching signals are followed most of the time but with occasional exploratory interruptions. Especially when the teaching signals have a small absolute value because then the system remains closer to the critical point which is formed by the branching point of the two fixed points corresponding to forward and backward motion. These interruptions cause the robot for example to move in curved fashion instead of driving in a straight line as the teaching signals dictate. The exploration around the teaching signals might be useful to find modes which are better predictable or more active, see Fig. 4.

Sensor teaching and distal learning

Because it is often easier to specify desired values for the sensor input than for the motor values, we will now show how to transfer the motor teaching paradigm to the sensory space. As a result the robot will be capable of performing imitation learning or can be trained based on a sensory trajectory that is recorded while it is passively moved around. Thus, a series of nominal sensations can be acquired that can serve as teaching signals. Providing the desired outputs in a different domain than the actual controller outputs leads to a distal learning problem (Jordan 1992; Stitt et al. 1994; Dongyong et al. 2000). Usually a forward model is learned that maps actions to sensations or more generally to the space of the desired output signals. Then the mismatch between a desired and occurred sensation can be backpropagated to obtain the required change of action. The backpropagation can also be done using an inversion of the forward model or by using a backward model, which learns the mapping from sensations to actions. In our case a forward model is already at hand, namely, it is given by the internal world model, see section Self-organized closed-loop control. Instead of a backpropagation we can also invert the world model directly since we use a linear implementation. The sensor teaching signals x_t^G are converted into motor teaching signals using the inverted model. Then one can apply the same mechanism as in Guidance by teaching.

Let us consider a more complicated example to illustrate the potential of this method, namely to induce in the Spherical robot a preference for rolling around one particular axis. For each axis the robot has an height sensor measuring the z-component of the vector attached to this axis, see Fig. 2b. A rotation of the robot around one of the internal axes is characterized by a zero sensor value for this axis while the remaining two sensor values oscillate periodically. In order to guide the robot into the rotation about, e.g., the first axis (shown in red in Fig. 2) we use a distal teaching signal whose first component is zero and the remaining two components contain the present sensor values such that the only learning signal relates to the first component.[2]

For an evaluation of the resulting behavior, we use the index of the internal axis with the largest rotational velocity. Figure 5 displays the percentage of times the first axis was actually the major axis of rotation. This information is given in dependence of the guidance factor γ (see Box 2). As expected there is no preferred axis of rotation without guidance ($\gamma = 0$). With distal learning the robot

[2] The teaching signal vector is given by $x_t^G = \left(0\, x_{t,2}\, x_{t,3}\right)^\top$, where $x_{t,i}$ are the sensor values at time t.

shows a significant preference for a rotation around the first axis up to 75%. For overly strong teaching, a large variance in the performance occurs. This is caused by a too strong influence of the teaching signal on the learning dynamics. If the the PSEF were ideally matched to the dynamics of the robot, arbitrary values of γ would be possible. In practice, however, the PSEF will provide only a hint toward the correct behavior and the details are left to be explored by the homeokinetic controller. At large values of γ the potential mismatches in the PSEF will influence to behavior of the robot in some runs which results in the large variance of the performance. Remember that the rolling modes can emerge due to the fine regulation of the sensorimotor loop to the working regime of the homeokinetic controller which cannot be maintained at a predominance of the PSEF.

Finally, one may ask why this method is unable to keep the controller at the desired rotational mode at all times? When the robot is in this rotational mode the teaching signal is negligible. However, the controller's drive to be sensitive will increase the impact of the first sensor, such that the mode becomes unstable again.

Guidance by cross-motor teaching

Finally we want to propose a guidance mechanism with internal teaching signals. As an example we want to influence the controller to prefer a mirror-symmetry in the motor patterns. This can be achieved by using the motor value of one motor as the teaching signal for another motor and vice versa. In other words, the teaching signals for each time step are given by a permutation of the motor values itself. This self-supervised teaching induced soft constraints which reduce the effective dimension of the sensorimotor dynamics and thus guide the self-organization along a sub-space of the original control problem.

Let us consider the TwoWheeled robot again and suppose the robot should move mostly straight, not get stuck at obstacles or in corners and cover substantial parts of its environment. We will see that all this can be achieved by a simple guidance of the homeokinetic controller where both motors are mutually teaching each other.[3]

For experimental evaluation we placed the robot in an environment cluttered with obstacles and performed many trials for different values of the guidance factor. In order to quantify the influence of the guidance we recorded the trajectory, the linear velocity, and the angular velocity of the robot. We expect an increase in linear velocity because the robot is to move straight instead of turning. For the same reason the angular velocity should go down. In Fig. 6

[3] The teaching signal is $y^{\mathrm{G}}_{t,1} = y_{t,2}$ and $y^{\mathrm{G}}_{t,2} = y_{t,1}$.

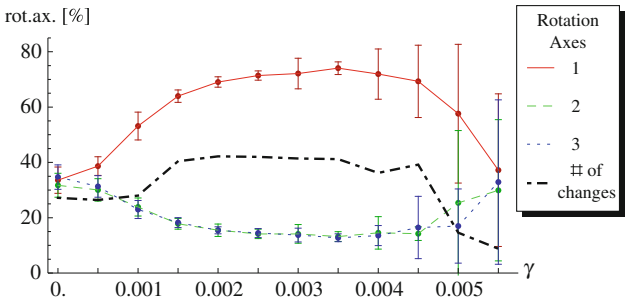

Fig. 5 Spherical robot and its behavior guided to rotate around the first internal axis. Behavior for the distal learning task. The figure shows the percentage of rotation around each of the internal axes and the number of times the behavior was changed for different values of the guidance factor γ (no teaching for $\gamma = 0$). The rotation around the first axis is clearly preferred for non-zero γ. The mean and standard deviation are plotted for 20 runs each 60 min long, excluding the first 10 min (initial transient, no guidance). For too large values of the guidance factor the self-organization process is too much disturbed such that the robot gets trapped in a random behavior (see *dashed-dotted line*). Parameters: $\epsilon_C = \epsilon_A = 0.1$, update rate 100 Hz

the behavioral quantification and sample trajectories are plotted. In addition, the relative area coverage[4] is shown, which reflects how much more area of the environment was covered by the robot with guidance compared to the case without. As expected, the robot shows a distinct decrease in mean turning velocity and a higher area coverage with increasing values of the guidance factor until the guidance dominates the behavior. Note that the robot is still performing turns and drives both backwards and forwards and does not get stuck at the walls, as seen in the trajectory in Fig. 6b. The properties of the homeokinetic controller, such as sensitivity and exploration, remain up to a certain strength of the guidance.

Discussion

We have presented here two new methods for guiding self-organizing behavior that are based on teaching signals. Desired motor patterns were specified by means of an error function that was integrated into the learning dynamics. The strength of guidance can be conveniently adjusted. Because teaching information is often given in the sensor space whereas learning is performed in the motor representation, a transformation is necessary which is obtained from the adaptive internal world model. The feasibility of both approaches was demonstrated by robotic experiments.

We introduced cross-motor teachings in order to be able to specify relations between different motor channels. If it

[4] The area coverage of the trajectory is calculated using a box-counting method.

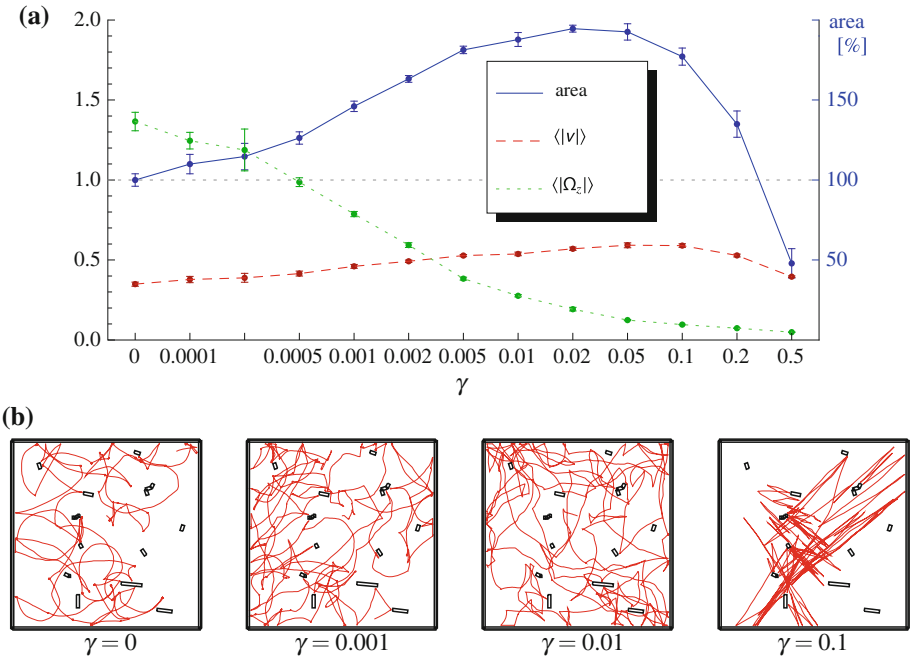

Fig. 6 Behavior of the TwoWheeled robot when guided to move preferably straight. **a** Mean and standard deviation (of 5 runs each 20 min long) of the area coverage,, the average velocity $\langle|\boldsymbol{v}|\rangle$, and the average turning velocity $\langle|\omega_z|\rangle$ for different values of the guidance factor γ. Area coverage (box counting method) is given relative to the case without influence ($\gamma = 0$: 100%) (right axis). The robot is driving straighter and its trajectory covers more area for larger γ, until at large γ the teaching dominates the behavior of the robot. **b** Example trajectories for different guidance factors. Parameters: $\epsilon_C = \epsilon_A = 0.01$, update rate 100 Hz

is known or desired that certain degrees of freedom of a robot should move in a coherent way, e.g., symmetrical or anti-symmetrical, then these relation can be injected as soft constraints that reduce the effective dimensionality of the system. As an example, the TwoWheeled robot showed that by enforcing the symmetry between the left and right wheel the behavior changes qualitatively to straight motion.

The exploratory character of the controller is nevertheless retained and helps to find a behavioral mode even if the specification of the motor couplings is partially contradictory. The resulting behaviors are not enforced by the algorithm. For example the TwoWheeled robot can choose freely between driving forward or backward whereas in direct teaching the direction of driving is obviously dictated by an external teacher. Furthermore, it is evident that the robot remains sensitive to small perturbations and continues to explore its environment.

Guided self-organization shares some properties with other approaches to autonomous robot control such as evolutionary algorithms (Nolfi et al. 2001) and reinforcement learning (RL) (Sutton et al. 1998). Evolutionary algorithms can optimize the parameters of the controller and are able to produce the same behaviors as we found in this study cf. (de Margerie et al. 2007; Ijspeert et al. 1999), if the fitness function is carefully crafted. More specifically, recent studies (Mazzapioda et al. 2009; Prokopenko et al. 2006) in evolutionary robotics have shown that the pursuit of task-independent aims can be helpful in optimizing a specific objective. This approach is related to the combination of (task-independent) self-organization and specific goals in this study. A critical experiment for the guided self-organization would investigate high-dimensional systems that cannot be decomposed into identical components.

A further difference is that self-organizing control is merely modulated by guidance, whereas evolutionary algorithms tend to converge to a static control structure. RL uses discrete actions or a parametric representation of the action space. In either case, high-dimensional systems will cause slow convergence. Recent findings with a chain-like robot (Martius and Herrmann 2011) show a clear advantage of cross-motor teaching in comparison to generic RL although similar relations among the actions in RL compensate part of this drawback. Natural actor-critics (Peters et al. 2005) may bring a further improvement of the RL control, but natural gradients can also be incorporated here. A decisive advantage of cross-motor teaching may be that goal-directed behaviors emerge within the self-organization of the dynamics from a symbolic description of the problem and do not need continuous training data such as in imitation learning (Peters 2008).

It is, however, clearly an interesting option to adapt cross-motor teaching to an imitation learning scenario. Although delayed rewards are still non-trivial for continuous domains, RL can cope with them in principle, while the guidance with rewards (Martius et al. 2007) requires instantaneous rewards.

Acknowledgments Both authors are grateful to Ralf Der for fruitful discussion. The project was supported by grants #01GQ0811 and #01GQ0432 within the National Bernstein Network Computational Neuroscience.

References

Abu-Mostafa YS (1995) Hints. Neural Comput 7(4):639–671

Amari S (1998) Natural gradients work efficiently in learning. Neural Comput 10(2):251–276

Cannon WB (1939) In: The wisdom of the body. Norton, New York

Choi J, Wehrspohn RB, Gösele U (2005) Mechanism of guided self-organization producing quasi-monodomain porous alumina. Electrochimica Acta 50(13):2591–2595

Der R (2001) Self-organized acquisition of situated behavior. Theory Biosci 120:179–187

Der R, Herrmann M, Liebscher R (2002) Homeokinetic approach to autonomous learning in mobile robots. In: Dillman R, Schraft RD, Wörn H (eds) Robotik 2002, no.1679 in VDI-Berichte. VDI, Berichte, pp 301–306

Der R, Hesse F, Martius G (2006) Rocking stamper and jumping snake from a dynamical system approach to artificial life. Adapt Behav 14(2):105–115

Der R, Liebscher R (2002) True autonomy from self-organized adaptivity. In: Proceedings of EPSRC/BBSRC International workshop on biologically inspired robotics. HP Labs, Bristol

Der R, Martius G, Hesse F, Güttler F (2009) Videos of self-organized behavior in autonomous robots. http://robot.informatik.uni-leipzig.de/videos

Di Paolo E (2003) Organismically-inspired robotics: homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In: Murase K, Asakura T (eds) Dynamical systems approach to embodiment and sociality, pp 19–42

Dongyong Y, Jingping J, Yuzo Y (2000) Distal supervised learning control and its application to CSTRsystems. In: SICE 2000. Proceedings of the 39th SICE annual conference, Iizuka, pp 209–214

Herrmann JM (2001) Dynamical systems for predictive control of autonomous robots. Theory Biosci 120:241–252

Ijspeert AJ, Hallam J, Willshaw D (1999) Evolving swimming controllers for a simulated lamprey with inspiration from neurobiology. Adapt Behav 7(2):151–172

Jordan MI, Rumelhart DE (1992) Forward models: Supervised learning with a distal teacher. Cognit Sci 16(3):307–354

Kelso JAS (1995) Dynamic patterns: the selforganization of brain and behavior. The MIT Press, Cambridge

de Margerie E, Mouret JB, Doncieux S, Meyer JA (2007) Artificial evolution of the morphology and kinematics in a flapping-wing mini UAV. Bioinspiration Biomim 2:65–82

Martius G, Herrmann JM (2011) Tipping the scales: guidance and intrinsically motivated behavior. In: Proceedings of advances in artificial life, 11th European Conference (ECAL 2011). MIT Press, pp 506–513

Martius G, Herrmann JM, Der R (2007) Guided self-organisation for autonomous robot development. In: Costa e FA (ed) Proceedings of advances in artificial life, 9th European Conference (ECAL 2007), LNCS, vol. 4648. Springer, San Francisco, pp 766–775

Mazzapioda M, Cangelosi A, Nolfi S (2009) Evolving morphology and control: a distributed approach. In: IEEE congress on evolutionary computation, New Orleans, pp. 2217–2224

Nolfi S, Floreano D (2001) Evolutionary robotics. The biology, intelligence, and technology of self-organizing machines. MIT Press, Cambridge

Peters J, Schaal S (2008) Natural actor-critic. Neurocomputing 71(7-9):1180–1190

Peters J, Vijayakumar S, Schaal S (2005) Natural actor-critic. In: Proceeidngs of the 16th European conference on machine learning (ECML 2005). Springer, Porto, pp. 280–291.

Prokopenko M (2008) Design vs self-organization. In: Prokopenko M (ed) Advances in applied self-organizing systems. Springer, London, pp. 3–17.

Prokopenko M (2009) Guided self-organization. HFSP J 3(5):287–289

Prokopenko M, Gerasimov V, Tanev I (2006) Evolving spatiotemporal coordination in a modular robotic system. In: Nolfi S, Baldassarre G, Calabretta R, Hallam JCT, Marocco D, Meyer JA, Miglino O, Parisi D (eds) SAB, LNCS, vol. 4095. Springer, Heidelberg, pp 558–569.

Rodriguez A (2007) Guided self-organizing particle systems for basic problem solving. Ph.D. thesis, University of Maryland, College Park

Stitt S, Zheng YF (1994) Distal learning applied to biped robots. In: Proceedings of the IEEE International Conference on robotics and automation. IEEE Computer Society, San Diego, pp 137–142

Sutton RS, Barton AG (1998) Reinforcement learning: past, present and future. SEAL, Florham Park, pp 195–197

Verschure PFMJ, Kröse BJA, Pfeifer R (1992) Distributed adaptive control: the self-organization of structured behavior. Robot Auton Sys 9(3):181–196