Berthouze, L., Prince, C. G., Littman, M., Kozima, H., and Balkenius, C. (2007).
Proceedings of the Seventh International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems. *Lund University Cognitive Studies, 135*.

Reflexes from Self-Organizing Control in Autonomous Robots

Frank Hesse* **Ralf Der**** **J. Michael Herrmann***
*Bernstein Center for Computational Neuroscience and Inst. f. Nichtlineare Dynamik
Universität Göttingen, Bunsenstr. 10, 37073 Göttingen, Germany

**Inst. f. Informatik, Universität Leipzig, Augustuspl. 10/11, 04109 Leipzig, Germany

Abstract

Homeokinetic learning provides a route to the self-organization of elementary behaviors in autonomous robots by establishing low-level sensomotoric loops. Strength and duration of the internal parameter changes which are caused by the homeokinetic adaptation provide a natural evaluation of external states, which can be used to incorporate information from additional sensory inputs and to extend the function of the low-level behavior to more general situations. We illustrate the approach by two examples, a mobile robot and a human-like hand which are driven by the same low-level scheme, but use the second-order information in different ways to achieve either risk avoidance and unconstrained movement or constrained movement. While the low-level adaptation follows a set of rigid learning rules, the second-order learning exerts a modulatory effect to the elementary behaviors and to the distribution of their inputs.

1. Introduction

Homeokinesis (Der et al., 1999, Der et al., 2002, Der et al., 2004) is a cybernetic approach to robot control. It is based on the dynamical systems approach, cf. e.g. (Tani, 2004). Homeokinesis represents a dynamical counterpart of the principle of homeostasis (Cannon, 1939, Ashby, 1954). According to the homeostatic principle, behavior results from the compensation of perturbations of an internal homeostatic equilibrium. Although this approach proves successful for the generation of simple systems (Di Paolo, 2003, Williams, 2004), it remains elusive how is scales up to situations, where, e.g., internal nutrient levels in an agent are to give rise to the specific form of a triggered behavior.

The homeokinetic principle, in contrast, provides a mechanism for the self-organization of a set of elementary movements, that are not directly caused by a triggering stimulus, but generated and executed in accordance to their controllability. Homeokinesis provides us with a mechanism to produce behaviors which does not depend on a reward signal or prescribed target state, but minimizes a unspecific internal error functional.

Homeostasis and homeokinesis are not alternatives, but should be seen as complementary principles. Since a single homeokinetic controller can generate a rich set of behaviors which are defined by intrinsic consistency, its output can be modulated or filtered such that a homeostatic principle is obeyed. Because homeokinetics reduces the manifold of behaviors to those which are controllable the complexity problem of the homeostatic approach to the generation of behavior is alleviated. It is also interesting to compare homeokinetic learning to reinforcement learning (Sutton and Barto, 1998). While reinforcement learning tries to maximize the expected reward for a sequence of single actions, homeokinetics selects coherent sequences which may later be used as building blocks in a higher-order reinforcement learning algorithm.

We present here an approach which is based on the interaction of sensory information processing and the control system. In this way self-organization of autonomous robots can be extended to more complex behaviors based on the interaction of the agent with its environment. We suggest an extension of the homeokinetic controller, where in addition to the minimization of the time-loop error also future errors are taken into account provided that they are predictable. It is not assumed that the predictability is based solely on the time series of measurements of the state, rather other sensory information is integrated. It is the predictability of the low-level learning process that triggers higher-order learning: If a situation requires low-level learning and repeats sufficiently often, then the higher-order learning mechanism contributes to avoid this situation. We will illustrate this two-layer architecture by the example of a robot that is able to change its internal parameters in order to escape from stalling in front of an obstacle and that subsequently learns to avoid the collisions by advancing the wall-related parameter changes to the time before the collision.

Low-level reflexes that are produced in the current model by a homeokinetic controller can be interpreted by high-level structures in different ways. We posit that a main goal of the interference by the high-level control consists in the modulation of the distribution of sensory inputs to the low-level control system. If low-level errors are interpreted as risky, the high-level control should succeed in avoiding situations where these errors occur, while in a different set-up errors form a challenge to the insufficient internal model of the agent that may learn more effectively if the frequency of errors is increased. We will consider both schemes in some detail while other examples are mentioned in passing. Before this, we will present a brief summery on the principle of homeokinesis in the next section. A learning rule is derived from this principle in Sect. 3. Sect. 4. describes the second-order learning. Experimental results in the learning architecture in realistic simulations are presented in Sections 5. and 6.

2. Homeokinesis

Starting from a standard approach to adaptive systems, we will discuss requirements for an objective function that allows a robot to acquire autonomously reproducible relations between actions and environmental feedback.

The behavior of the robot is governed by a controller K that generates motoric outputs

$$y = K(x; c) \tag{1}$$

as a function of the vector of sensory inputs $x = \{x_1, \ldots, x_d\}$. The effect of the controller depends on a parameter vector $c = \{c_i, \ldots, c_n\}$. For example, the c_i may represent weights of the input channels from which K calculates a squashed sum for each output. We further assume that the inputs are processed on a time scale that is short compared to the cycle time. Adaptivity of the controller is achieved based on an objective function E that contains possibly implicit information about desired behavior. In this sense it is sufficient to define E based on the probability of survival or, as we will proceed here, by a functional that is evaluated continuously by the robot itself.

Interaction of an agent with its environment includes sensitivity to sensory stimuli. The actions of the robot should appear to be caused by the inputs, although the particular form of the relation between inputs and outputs may change continuously due to the adaptation of the internal parameters. Whether or not a reaction of an robot is due to a particular stimulus cannot be decided by reinitialization in a similar perceptual situation if the robot is to learn autonomously. We therefore equip the robot with an internal model that enables the robot to compare a situation with a similar one that was encountered earlier. If the robot's objective were solely the reproducibility of a reaction then the robot would tend to run into trivial behaviors. This could mean e.g. that the robot behaves such that the inputs stay constant, which at least does not allow us to observe the robot's sensitivity. We focus therefore on the unavoidable differences between inputs and the corresponding predictions by the internal model. If these differences are small then the robot is obviously able to predict the consequences of its actions. If, in addition, the differences tend to increase due to new stimuli then the robot is still sensitive. Note that the differences can be decreased also by improving the internal model.

The main idea of the approach derives from the fact that a destabilization in time is dynamically identical to a stabilization backward in time. Because predictability in a closed loop enforces stability, a virtual inversion of time is the road to a destabilization without the loss of predictability. This idea will become more clear in the following formal description.

As a first step we introduce a virtual sensor value $\hat{x}$ by

$$\hat{x}_t = \arg\min_x \|x_{t+1} - \psi(x)\|, \tag{2}$$

which is optimal with respect to the prediction of the following input x_{t+1}, although generally different from the real input x_t. The predictor ψ is realized by a parametric function, e.g. an artificial neural network, which receives x and the controller parameters c as inputs and generates an estimate of the subsequent input.

We can interprete the calculation of $\hat{x}_t$ as a mechanism for editing an earlier input which is invoked once x_{t+1} is available. In order to minimize the effect of the editing, one should require that

$$\|x_t - \hat{x}_t\| \to \min, \tag{3}$$

which actually turns out to be the central criterion of the approach. Eq. 2 is nothing but a regularized solution of the possibly ill-posed equation

$$\hat{x}_t = \psi^{-1}(x_{t+1}), \tag{4}$$

which reveals that in principle an inverse loop function is used in order to produce $\hat{x}_t$, see also Fig. 1. In this sense we consider the sensory dynamics in inverted time while avoiding a conflict with causality.

Often instead of (3) the problem

$$\|x_{t+1} - \psi(x_t)\| \to \min \tag{5}$$

is considered which measures the forward prediction error. Eqs. 3 and (5) have in common that both criteria try to optimize the predictability of sensory inputs by the predictor ψ. The dynamical effects are, however, quite different. (5) causes the controller to decrease the distance of any noisy trajectory from the predicted value. This leads generically to a convergence of nearby trajectories, i.e. tends to stabilize the current behavior. Condition (3), in contrast, causes nearby trajectories to diverge. In (3) we shall use the abbreviation $v_t = x_t - \hat{x}_t$ in order to denote the predictor-based sensory shift. It is used in the definition of an energy function

$$E = \|v\|^2 . \qquad (6)$$

Because minimizing E minimizes the sensitivity of ψ^{-1}, cf. Eq. 4, the sensitivity of the function ψ with respect to variations of its arguments is maximized. The shift v is small if both $\xi = x_{t+1} - \psi(x_t)$ is small and the derivative of ψ is large. Hence, the two goals of sensitivity and predictability are implicit in (6). This becomes more obvious when setting $v = L^{-1}\xi$ with $L_{ij} = \frac{\partial}{\partial x_j}\psi_i(x)$. being the Jacobian of the system. The energy function is then

$$E = \left(L^{-1}\xi\right)^T \left(L^{-1}\xi\right), \qquad (7)$$

where it can directly be seen, that by gradient descent on E the modeling error ξ is decreased whereas the sensitivity of the system is increased by increasing the Jacobian L. However, an increase in sensitivity will tend to lower predictability and vice versa, such that the actual behavior can be expected to oscillate between periods of exploration and stabilization in a way which reflects the quality of the predictor and the complexity of the environment.

3. Learning rules for control

Because the energy (7) depends via the behavior of the robot also on the controller parameters (1), adaptive parameter changes can be achieved by a gradient flow on $E(x_t, c_t)$, where x_t denotes the trajectory of the robot in the sensory state space and c_t the current values of the controller parameters. The combined dynamics

$$x_{t+1} = \psi(x_t, c_t) + \xi_t \qquad (8)$$

$$c_{t+1} = c_t - \varepsilon_c \frac{\partial}{\partial c} E(x_t, c_t) \qquad (9)$$

describes both the effects of the environment and the controller on the sensory state of the agent as well as the adaptation of the internal parameters. The resulting state (8) and parameter dynamics (9) run concomitantly and form a dynamical system in the product space formed by x and c. Learning, in this sense, means to identify both a set of parameters as well as a region in x-space which are optimal with respect to E. It is possible that the learning process results in a limit cycle involving both parameters and states, or it may be even open-ended by allowing a robot to gradually explore a virtually unbounded environment.

The intrinsic parameters of the predictor ψ internal model of the environment, cf. Fig. 1, predicting how sensor values x are influenced by controller outputs y. ψ is adapted based on the prediction error $\|x_{t+1} - \psi(x_t)\|$ with learning rate ε_p. The ratio of ε_c in (9) and ε_p is crucial for the learning process. If $\varepsilon_p \approx \varepsilon_c$ the model is in principle able to track the changes in the behavior of the agent, while for $\varepsilon_p \ll \varepsilon_c$ the model is rather accumulating information about the environment. Note that ε_c corresponds to a time scale comparable to that implicit in the sensory dynamics in order to allow for rapid behavioral adaptation.

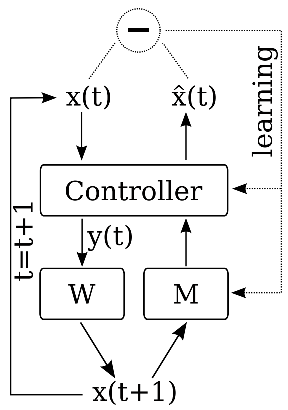

Figure 1: Sketch of the homeokinetic control scheme based on a sensorimotor loop. Sensor values $x(t)$ are used by the controller to generate motor commands $y(t)$ which are executed in the world (W). Subsequently, new sensor values $x(t+1)$ become available. A world model, denoted by M, which realizes a function $\psi(x(t)) \approx x(t+1)$, is simultaneously adapted. The goal of the parameter adaptation is to minimize the difference between the virtual and the true sensor value $x(t)$.

Ignoring the effects of nonlinearity in Eqs. 8 and 9 we find that the state in (8) is close to the largest eigenvector of L. Therefore, a learning rule based on Eq. 5 reduces the maximal eigenvalue of L. A learning rule based on (6), (7) will instead tend to increase the minimal eigenvalue of L by minimizing the maximal eigenvalue of L^{-1}. In this way more and more modes become activated and, if the noise ξ does not increase, the behavior becomes sensitive with respect to many different stimuli.

In the following we will use a linear controller $y = K(cx_t + h)$ for which Eq. 9 becomes

$$\Delta c = \mu a - 2\mu x (z - h) \qquad (10)$$

$$\Delta h = -2\mu (z - h) \qquad (11)$$

where a is the linear response of the environment to the action y and μ is a learning rate.

4. Hebbian second-order learning

The homeokinetic controller (1, 10-11) generates simple reactive behaviors that are interesting because of their flexibility. We are now going to modify the controller such that in addition prospective information can be exploited in order to generate preventive action, thereby relying on other information which may be available from more complex sensors and predictors. We propose to interprete such information in terms of the low-level control which may be advantageous if no background information can be referred to for the interpretation of the high-level information.

We extend the homeokinetic controller by an additional learning mechanism which brings about the avoidance of situations that cause a large modeling error before they occur. For this purpose, a mechanism is required that is able to predict the modeling error. The error function (7), minimized by the homeokinetic control layer, is extended by an additive contribution from context sensors which can be represented by

$$E = \left(L^{-1}(\xi + \zeta)\right)^T \left(L^{-1}(\xi + \zeta)\right) \qquad (12)$$

Here we introduced the prediction ζ of the error of the state estimation which is provided by a different input channel. This additional error will be associated to the low-level learning process by an additional Hebbian layer, cf. Fig. 2. In addition to the minimization of the state estimation error ξ the robot minimizes the prediction ζ of the state estimation error. ζ is defined to be small when context information is unavailable such that in these cases the actual behavior is produced by the low-level controller. Otherwise the behavior will be changed such that ζ is reduced. The extended energy function (12) applies only in the update rule of the threshold h, because the exploratory mode is characterized by a stationary non-zero c value.

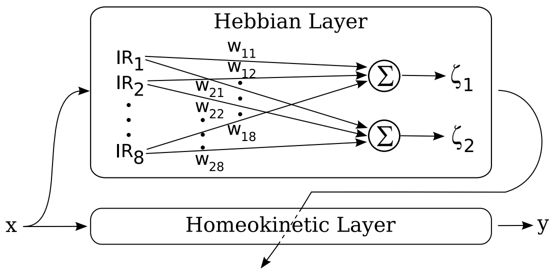

Figure 2: Sketch of the extended control structure. A Hebbian layer affects the homeokinetic controller by the predicted modeling error ζ which complements the low-level time-loop error.

The Hebbian layer is realized by a leaky integrator neuron with a linear output function for each of the sensor x, i.e. for each each sensor value x_i a predicted sensor value $\hat{x}_i$ and a specific modeling error ξ_i is used. All sensory information x^H is used as input to each neuron and weighted by the synaptic strength w_{ij} according to

$$\zeta_i = \sum_{j=1}^{m} w_{ij} x_j^H , \quad i = 1..n, \qquad (13)$$

with m being the number of sensors available to the Hebbian layer and n number of sensors available to the homeokinetic layer. The update rule for the weights is

$$\Delta w_{ij} = \varepsilon \xi_i x_j^H (1 - w_{ij}^2) \qquad (14)$$

where ε is a learning rate and $\xi_i x_j^H$ realizes Hebbian learning between the modeling error ξ_i of the homeokinetic layer and the input x_j^H of the Hebbian layer. A decay term $(1 - w_{ij}^2)$ is added, which restrains the weights from unlimited increase. Nevertheless, a sensory input of 1 weighted with a synaptic strength of nearly 1 would result in a predicted modeling error of about 1, which can realize an immediate change of the actual behavior as intended. By adding the predicted modeling error via the Hebbian layer to the actual modeling error, the homeokinetic controller can thus avoid situations which lack low-level predictability.

5. Foraging in a wheeled robot

In the general case we have a vector of sensor values $x_t \in \mathbb{R}^d$ at the discrete instants of time $t = 0, 1, 2, \ldots$. By way of example we may consider a two-wheeled robot, cf. Fig. 3, where the low-level controller receives the measured wheel velocities as input. In addition infrared sensors are available as context sensors.

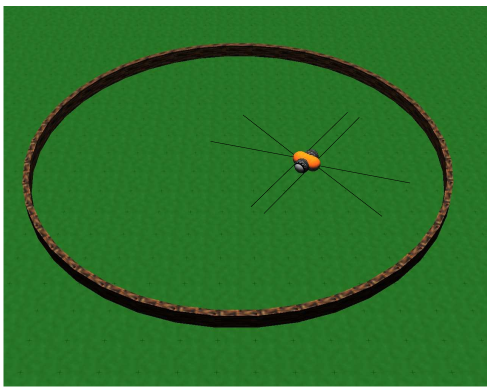

Figure 3: Experiments are performed with a two-wheeled robot in a cylindrical box. The robot is equipped with wheel counters and eight infrared sensors. The black lines indicate the IR sensor orientation and range. The sensor range is 3 and the diameter of the box is 14 length units.

The modeling error describes differences between predicted and measured wheel velocities. The predicted modeling error is used to modulate the homeokinetic layer in order to change the actual behavior before arriving at situations with a large modeling error, which refers to collision situations in the example.

In the experiments we will show that obstacle avoidance behavior of a two-wheeled robot equipped with infrared sensors can be obtained, solely based on the intrinsic properties of the system. The effectiveness of the obstacle avoidance is not perfect since the system tries occasionally to explore also the regions near the boundaries. Nevertheless the time the robot spends near obstacles is drastically reduced, cf. Fig. 5.

The initial setup of the experiments consists of a simulated two-wheeled robot with infrared sensors, placed in a circular box with radius 7, see Fig. 3.

The Hebbian layer is provided with proximity information from eight infrared sensors with a sensor range of three length units. In order to suppress small noisy activity in the infrared sensors, only sensor values larger than 0.15 are considered. The synaptic strength w_{ij} of the Hebbian layer are initialized with zeros. The parameters of the homeokinetic layer are initialized with small random values.

In a first experiment only the homeokinetic layer was used. Experiment 2 was done using the extended controller. Each experiment runs for one hour simulated real time in a simulation environment that is designed to realistically reproduce physics. To obtain some information about long-term stability a third experiment was conducted that lasted 24 hours.

The trajectory of the robot in the two experiments is plotted in Fig. 4. The positions of the robot concentrate increasingly to the inner obstacle-free region when using the Hebbian control layer as compared to pure homeokinetic control. The histogram of the robots distance from the center of the box illustrates the effect of the learning scheme, see Fig. 5. During the first part of the experiment (top row) the Hebbian layer started to adapt but shows hardly any effect on the robots behavior yet. Hence the histograms show similar distributions.

The bottom row of Fig. 5 shows histograms of the robots position during a later part of the experiment where the influence of the Hebbian control layer is dominant. Without access to the Hebbian layer the robots probability of staying near the wall is approximately three times higher than being at any other distance from the center, cf. Fig. 5 (bottom left). This is caused by the fact that in the central obstacle free region of the box behaviors are more stable due to the small modeling error and hence larger distances are covered by the robot. Whereas in the region near the wall behaviors change more often due to a larger modeling error and the robot is not able to cover large distances. Therefore the robots probability to stay near the wall is higher. When enabling the Hebbian layer the robots probability of being near the wall is drastically reduced and the highest probability is now shifted towards the center of the box, see Fig. 5 (bottom right).

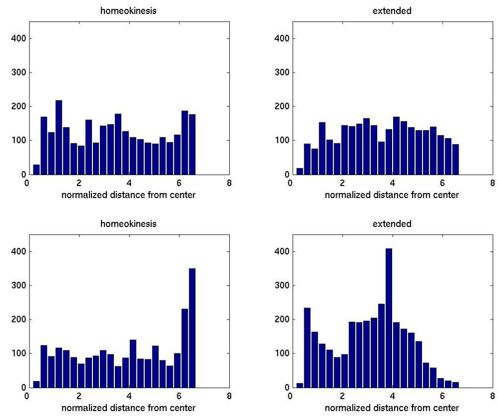

Figure 5: Histogram of the robots distance from center normalized by the resp. areas for pure homeokinetic control (left column) and the extended controller (right column) of the first 15 minutes (top row) and the last 15 minutes (bottom row) of the experiments with a total time of 1 hour. In the initial phase the Hebbian layer is not yet functional and both controller show comparable results. In the later part of the experiment (bottom row) the mean occupancy has shifted away from the wall towards the center of the box in the case of the extended controller.

The predicted modeling error of the Hebbian layer leads to a change of the actual robot behavior before the collision region is reached. Since the selection of the following behavior is not constrained the robot can still reach the collision area, but with much less probability. This can be interpreted as a flexibility of the system which continues to explore the collision area.

The usage of the predicted modeling error in the homeokinetic layer leads to pre-collision changes of the robots behavior rather than to the trivial solution where the robot stops somewhere in the central region of the box. In Fig. 6 the traveled distance of the robot with and without usage of the Hebbian layer is shown. Regions of inactivity are essentially absent. Also the total traveled distance is not reduced by incorporating the Hebbian layer. In the 24-h experiments no stability problems of the system were observed.

The weights of the Hebbian layer during the 24-h experiments show that the learned correlations indeed have an effect on the behavior of the robot,

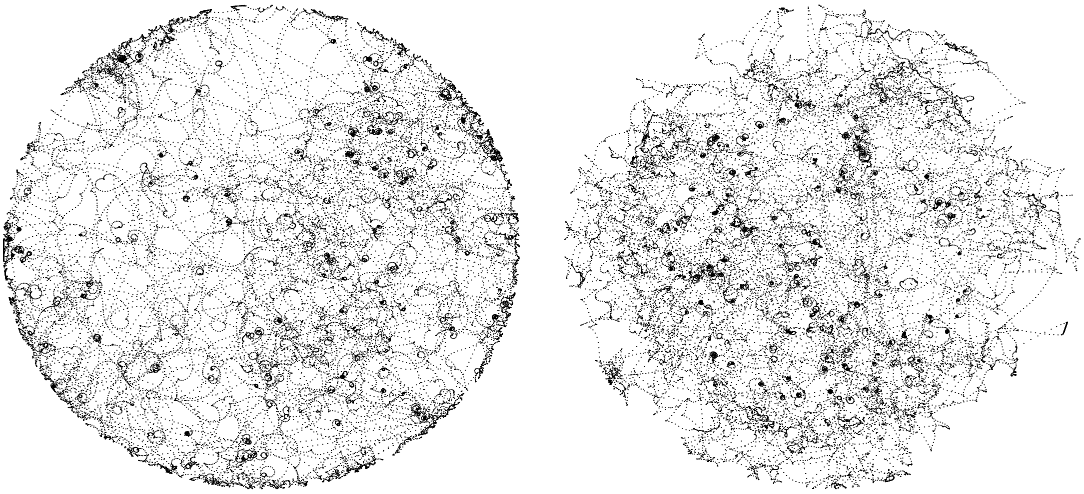

Figure 4: Trajectory of the robot using pure homeokinetic control (left) and the extended controller (right). An increasing concentration of the robots positions in the inner obstacle free part of the cylindrical box can be identified when using the Hebbian control layer, as compared to pure homeokinetic control.

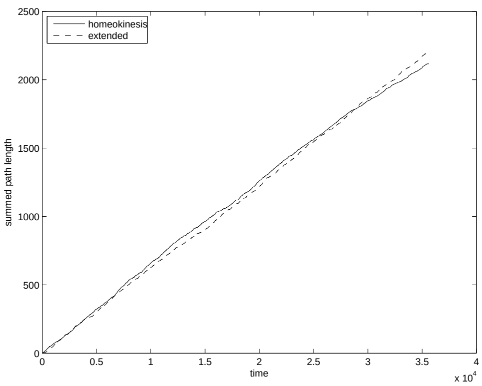

Figure 6: Cumulative distance traveled by the robot over time using pure homeokinetic control and the extended controller. The traveled distances in the two experiments are comparable, indicating that the Hebbian layer did not reduce the activity of the robot.

cf. Figs. 7. The two front infrared sensors are included with negative sign. Note that, if the wheel counters indicate forward motion by $x > 0$ then the predicted velocity will typically also be positive $\hat{x} > 0$. Near a wall the front infrared sensor will be active, but after collision the velocity sensor will yield $x = 0$, while the prediction is still $\hat{x} > 0$. Hence, ξ will be negative. So by converging to negative weights for the front infrared sensors the Hebbian layer extracts this correlation and is able to predict a negative future modeling error ζ. The same holds true for the rear infrared sensors with inverted sign for weights and modeling error. For the sideward sensors the correlations are not significant.

If the effect of the additional error term is inverted the robot will move only in the vicinity of the wall. In this way the robot increases its opportunity to adjust its internal parameters such that it is able to move freely near walls. The robot's preference for wall in this modified scheme suggest it as a model for a foraging rat, cf. e.g. (Tamosiunaite et al., 2008). We will study the modified principle in a more complex hardware set-up in the following section.

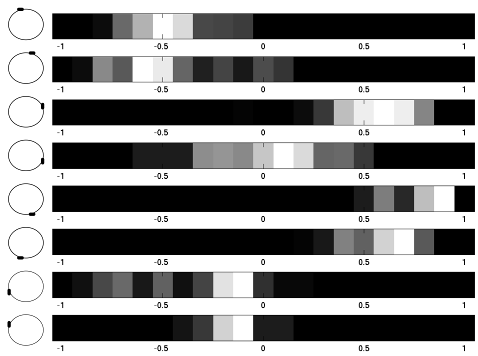

Figure 7: Histogram of the weights of the Hebbian layer contributing to ζ_1 for a long-term experiment (24 h real time) with extended controller. The labels at the y-axis correspond to the eight infrared sensors. Front and rear sensor weights have negative and positive sign, resp., indicating the ability of the Hebbian layer to correctly extract the correlations between modeling error ξ and IR sensor activity. For details see text.

6. Gripping in a human-hand model

For the further evaluation of the context-based exploration we programmed a model of a human hand with 5 degrees of freedom, see Fig. 8. All joints are controlled by bidirectional motors that mimic the interplay between flexor and extensor muscles. The effect of a motor action is measured by motion sensors, which serve as input to the low-level homeokinetic controller. Each finger is controlled by an individual controller such that interactions between the fingers are possible only via the environment. If no object is present for manipulation the finger become quickly engaged in vivid movements which can be interpreted as an exploration of the dynamical range. In the presence of an object the modeling errors increase considerable when the fingers touch the object, because this is not predicted by the internal model. Context information about objects in the hand is provided by infrared sensors in the finger tips.

In this experiment we exploit the directionality of the Hebbian weights, see Fig. 7, by directly adding the output of the higher layer to the update of the threshold h in (11). This will give the same result as applying the scheme of the previous section. The fingers will flinch when arriving close to the surface of the object but remain active otherwise like in the free case. By changing the sign of the contribution of the higher layer to the bias update we shape the behavior of the system in order to show a gripping reflex. Results presented in Figs. 8 and 9 show that soon, after an object is presented, a grip at the object is realized due to the domination of the Hebbian layer. If the object is removed and presented again the hand closes and the fingers grab the object.

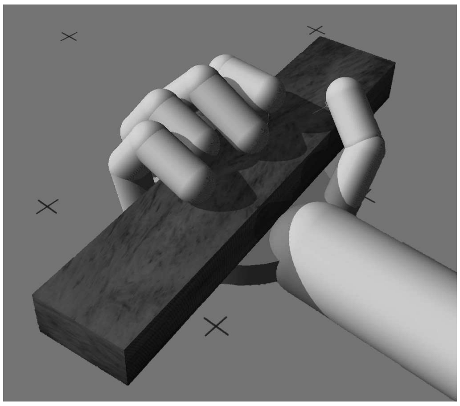

Figure 8: Simulation of a human hand with multiple degrees of freedom. The hand is equipped with motion sensors at all joins and infrared sensors at the finger tips. It is operated in a fully exploratory mode with or without a manipulated object.

7. Conclusion

In the experiments realistically simulated hardware agents acquired low-level behaviors which are characterized by simultaneous sensitivity and coherency. The basic behaviors are obtained from the interplay of a mildly destabilizing controller with the environment which is constrained by the prediction quality achieved by an internal model. In unforeseen situations, i.e. "obstacles", parameter changes are triggered which are time-consuming and may even cause unlearning of previously acquired behaviors. The proposed second-order learning schemes are coping with such a situation in different ways (Berthouze and Lungarella, 2004): Either the robot is controlled such as to avoid these situation which generates an interpretation of additional sensory inputs in terms of the low-level affordances. Or the robot is guided towards these situations in order to further improve its prediction quality. The decision which mode of operation of the second-order learning is to be activated is to be taken in dependence of the quality of the internal model such that increasing prediction quality should favor the explorative mode while unsurmountable errors should lead to a preference of the avoidance behavior. The explorative character of the low-level self-organizing con-

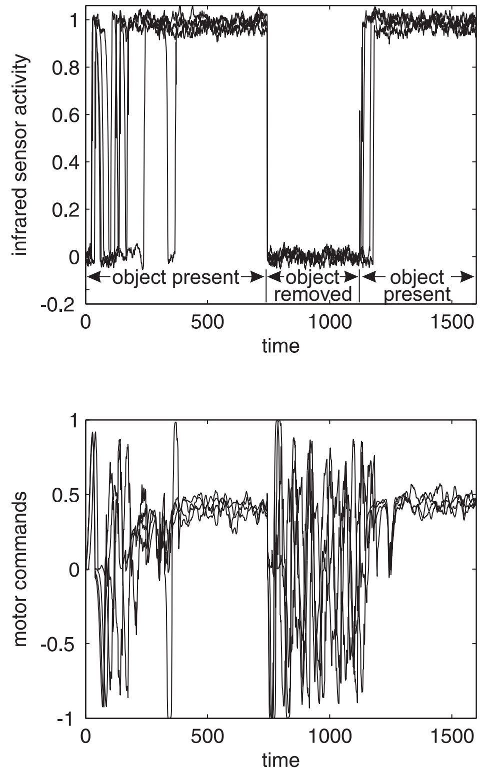

Figure 9: The finger movements that are initiated by the self-organizing controller soon converge to a grip at the object (high infrared sensor activity) with only small deviation of single fingers from the surface. When the object is removed the exploratory movements restart. If the object is present the fingers will grip it again since the Hebbian layer already learned this reflex.

troller is retained in both cases and the robot still explores occasionally risky regions and is hence able to adapt to slow changes in the environment. The work shows also parallels to the early motor development in biology, cf. e.g. (Kuniyoshi and Sangawa, 2006), and provides a scheme for the formation of reflexes based on an approach to the self-organization of autonomous behavior.

Acknowledgment

This work was supported by the BMBF in the framework of the Bernstein Centers for Computational Neuroscience, grant number 01GQ0432. Discussions with C. Kolodziejski and G. Martius are gratefully acknowledged.

References

Ashby, W. R. (1954). *Design for a Brain*. Chapman and Hill, London.

Berthouze, L. and Lungarella, M. (2004). Motor skill acquisition under environmental perturbations: on the necessity of alternate freezing and freeing. *Adaptive Behavior*, 12:47–631.

Cannon, W. B. (1939). *The Wisdom of the Body*. Norton, New York.

Der, R., Herrmann, M., and Liebscher, R. (2002). Homeokinetic approach to autonomous learning in mobile robots. In Dillman, R., Schraft, R. D., and Wörn, H., (Eds.), *Robotik 2002*, number 1679 in VDI-Berichte, pages 301–306. VDI.

Der, R., Hesse, F., and Liebscher, R. (2004). Self-organized exploration and automatic sensor integration from the homeokinetic principle. In *Proc. SOAVE Ilmenau 04*, Ilmenau.

Der, R., Steinmetz, U., and Pasemann, F. (1999). Homeokinesis - a new principle to back up evolution with learning. In *Computational Intelligence for Modelling, Control, and Automation*, volume 55 of *Concurrent Systems Engineering Series*, pages 43–47, Amsterdam. IOS Press.

Di Paolo, E. (2003). Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In Murase, K. and Asakura, T., (Eds.), *Dynamical Systems Approach to Embodiment and Sociality*, pages 19 – 42.

Kuniyoshi, Y. and Sangawa, S. (2006). Early motor development from partially ordered neural-body dynamics: experiments with a cortico-spinal-musculo-skeletal model. *Biological Cybernetics*, 95(6):589–605.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press/Bradford Books.

Tamosiunaite, M., Ainge, J., Kulvicius, T., Porr, B., Dudchenko, P., and Wörgötter, F. (2008). Path-finding in real and simulated rats: On the usefulness of forgetting and frustration for navigation learning. *J. Comp. Nsci.*, submitted.

Tani, J. (2004). Symbols and dynamics in embodied cognition: Revisiting a robot experiment. In Butz, M. V., Sigaud, O., and Gerard, P., (Eds.), *Anticipatory Behavior in Adaptive Learning Systems*, pages 167–178. Springer-Verlag.

Williams, H. (2004). Homeostatic plasticity in recurrent neural networks. In Schaal, S. and Ispeert, A., (Eds.), *From Animals to Animats: Proceedings of the 8th Intl. Conf. On Simulation of Adaptive Behavior*, volume 8 of *8*, Cambridge MA. MIT Press.