Self-organized exploration and automatic sensor integration from the homeokinetic principle

Ralf Der, Frank Hesse and René Liebscher
Universität Leipzig, Institut für Informatik
POB 920
D-04009 Leipzig
{der|fhesse|liebschr}@informatik.uni-leipzig.de

Abstract

Starting from the homeokinetic principle introduced earlier the present paper presents simple learning rules for the neurons of a closed loop robot controller. These learning rules are shown in a simple application case to realize a self-learning autonomous robot which can survive in a sufficiently simple world without any further external help. In particular we demonstrate that sensors are automatically integrated according to their response strength as soon as they deliver a signal to the controller. Moreover the system also can deal with the problem of a rapid change in the properties of the sensors. The basic effect observed is that the learning rule drives the robot in the sensorimotor loop into an explorative mode of behavior which however is sensitive to the reactions of the environment by way of the model error. From a dynamic systems point of view we have a closed loop control system with a pitchfork or a Hopf bifurcation (if the learning of the threshold is included) and the effect of the learning is to drive the system to a regime slightly above the bifurcation point where such systems are known to be particularly sensitive.

1 Introduction

The central interest of our work is the self-organized acquisition of behaviors for autonomous robots. Our work is based on the belief that true autonomy must involve the phenomenon of emergence. Before giving some ideas how this could be realized in the robotic domain let us first illustrate the goal in a realistic case. Consider a robot with a neural network controller with synaptic weights initially in the *tabula rasa* condition. So there is no reaction of the robot to its sensor values and activities, if present at all, are only stochastic ones. The robot is to be in an environment with static and possibly also dynamic objects. The task now is to find an objective function for the adaptation of the controller which is entirely internal to the robot driving the parameters so that the robot will start acting and while acting to explore and develop its perception of the world and of object related behavior.

Our work aims at finding general principles for the realization of this program. One of our approaches is given the name of homeokinesis [4], [5], [2] which is the dynamical pendant of homeostasis as introduced by Cannon [1] and later Ashby [7] and in the embodied intelligence approach

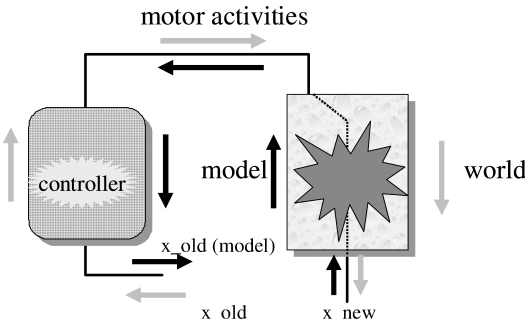

Figure 1. Dynamic harmony is reached if the agent manages to balance the information streams forward and backward in time. The time loop error is given by the difference between the true sensor values x_{old} and the reconstructed $x_{recon} = Model^{-1}(x_{new})$ ones.

[6] as discussed in [2]. We may call this principle also the principle of dynamic harmony between internal and external world. This principle can be given a constructive formulation in the following way. We consider the case of an autonomous agent and provide the agent with an adaptive model of its behavior. A learning signal for both the model and the controller is derived from the misfit between the real behavior of the agent in the world and that predicted by the model. As we could show with several examples this misfit is minimized if the agent exhibits a smooth, controlled behavior.

In this way, a learning signal for the adaptation of the behavior is derived from a purely internal perspective. However using a predictive model will lead to active behavior modes only if the agent is given a drive for activity (or curiosity or the like) from outside. We therefore use a retrospective model where the modeling step is backward in time. The succession of steps forward through the world and backward through the model can be considered as running through a time loop. The aim of adaptation now is the minimization of the time loop error. We can also say that the dynamical harmony is reached if the information streams forward and backward in time are balanced, see Figure 1.

The basic effect of this paradigm is understood in the following way. The model error is propagated backward through time to the input of the controller. Decreasing the time loop error means that when looking backward in time the behavior is to damp the influence of noise (model error). However in the physical (forward) direction of the time arrow this means that the behavior is to magnify the noise. As a consequence the overall feed-back strength of the sensorimotor (SM) loop and hence the response of the robot to the sensors is increasing until the dynamics starts exploding. Then the nonlinearities of the world and the controller will confine the further increase of the activity. As a result the system is driven towards a bistable regime (or a limit cycle behavior, see below) with smooth switching between behaviors (directions of motor velocities) the frequency of which is dominated by the strength of the model error.

The principle can be translated into concrete learning rules for the synapses of the controller neurons. In the present paper we want to investigate these learning rules in a simple case and show that it generates on the one hand an explorative behavior of the robot which is sensitive to the environment (the frequency effect) and on the other hand we demonstrate that sensors are automatically integrated.

2 Essentials of the approach

Essential features of our self-learning approach to behavior control can already be seen by considering a very simple system consisting of a neuron controlling the velocity of a robot.

2.1 The sensorimotor loop

Let us consider a closed loop velocity control of a robot corresponding to a sensorimotor (SM) loop which is closed via the wheels only. Our controller consists of a single leaky-integrator neuron with membrane potential z updated in the time step as

$$\tau^{-1}\Delta z_t = -z_t + cx_{t+1} + H \tag{1}$$

where the input is the true wheel velocity x as measured by the wheel counter and H being a bias (threshold). The update is carried out when the new sensor value x_{t+1} arrives. The output of the neuron

$$y_t = \tanh(z_t) \tag{2}$$

is the target wheel velocity of the robot. We make a simple model assumption relating the expected true velocity $\hat{x}_{t+1}$ to the target velocity y_t as

$$\hat{x}_{t+1} = ay_t \tag{3}$$

where the constant a (the response strength of the channel) can be learned by minimizing the error

$$E = (x_{t+1} - ay_t)^2 \tag{4}$$

for samples (x_{t+1}, y_t) obtained on-line in each time step $t = 0, 1, \ldots$.

Our SM loop is described by the system

$$x_{t+1} = ay_t + \xi_t \tag{5}$$

where ξ_t accounts for the differences between the true wheel velocity and the one predicted by the deterministic model.

The behavior of the robot essentially depends on the values of c and H. This is seen best by observing that by means of eq. 3 we get the closed dynamics for the membrane potential as

$$\tau^{-1}\Delta z_t = -z_t + ca\tanh(z_t) + H + c\xi_t \tag{6}$$

The essential point of this dynamics is already seen from a linear stability analysis obtained by replacing $\tanh z$ with z (for small z) so that with $K = ca$

$$\tau^{-1}\Delta z_t = -(1-K)z_t + c\xi_t \tag{7}$$

(case $H = 0$ for simplicity). The deterministic system is seen to have a stable fixed point (FP) $z^* = 0$ as long as $0 < K < 1$. This means that z_t is fluctuating around zero (z_t is the time-discrete version of the Ornstein-Uhlenbeck process) and the robot essentially executes a random walk. At $K = 1$ we have a pitchfork bifurcation so that for $K > 1$ the FP $z^* = 0$ is destabilized and there are two new FPs at $z = \pm z^*(c)$. With the noise included we get a stochastic bifurcation at some value $K_c > 1$ which depends on the strength of the noise. If $K > K_c$ the robot moves either forward or backward with constant velocity (apart from fluctuations) and with K slightly above K_c the noise can switch the system between these two alternatives.

One of the benefits of this closed loop control system consists in the following. When colliding with an obstacle the wheels are blocked, so that $x = 0$ and z decays. If there is some (small) additional noise in the dynamics of the membrane potential z it will fluctuate around zero. These fluctuations are amplified if they are of the right sign, i.e. if the robot is moving away form the obstacle. Hence after a short time the robot is found to move away from the obstacle. We may say that in this elementary sense the robot is able to survive.

In the case of finite H the FPs and hence the velocity of the robot are obtained from

$$z = K \tanh(z) + H$$

so that they are a function of both H and c and we may use these parameters in order to define the behavior of the robot. Our ambition now is to find adaptation rules for these parameters so that the robot determines its behavior independently.

2.2 The parameter dynamics

The homeokinetic principle (gradient descending the time loop error) in some approximation produces the following dynamics for the synaptic strength c and the bias H

$$\begin{aligned} \Delta c &= \mu a - 2\mu z x \\ \Delta H &= -2\mu z \end{aligned} \tag{8}$$

where

$$\mu = \varepsilon \xi^2 g'(z)$$

is a modified update (learning) rate, and $g' = \tanh'(z) = 1 - \tanh^2(z)$. The parameter dynamics is to be used concomitantly with the z dynamics so that the parameters c and H in eqs. 1 or 6 are now time dependent. As we will see below the time scale for the change of in particular H is on the level of the behavior so that in other words the behavior is essentially controlled by the dynamics of H. This is different from the usual paradigm of learning where we have a learning and a performance phase or where there is a separation of time scales for learning and behaving.

For a discussion we consider the case $H = 0$ first. The dynamics for c, eq. 8, consists of the driving term μa and an anti-Hebbian term given by the product of the input into the synapse times the membrane potential of the neuron both quantities being felt directly at the synapse. In order to discuss the effects of the two terms we assume that we start the system with $K < 1$ (in the tabula rasa condition, i.e. $c = K = 0$, e.g.) so that z fluctuates around zero. Hence the anti-Hebbian term is negligible and the driving term is seen to increase the value of K since $\Delta K = \Delta(ca) = \mu a^2$. Once $K > 1$ is reached the velocity increases exponentially so that the robot starts moving. With $y^2 > 0$ the anti-Hebbian term comes into play and the increase of c is stopped if $a = 2zx$ hence $1 = 2z \tanh z$ or $1 = 2Ky^2$ which happens at $K \approx 1.2$ where $y \approx \pm 0.65$. The direction of the robot (sign of the velocity) is arising from a spontaneous breaking of the $x \to -x$ symmetry inherent in the complete (i.e. parameter and state) dynamics. Note that K is the feed-back strength in the SM loop so that we observe a self-regulation of the system to a feed-back strength which is slightly supercritical.

With $H \neq 0$ the FP of course depends also on the value of H and there is a hysteresis effect w.r.t the change of H which is the larger the larger K. Under the parameter dynamics the value of H changes and it is easily seen that a limit cycle behavior is obtained. The value of c is seen to slightly oscillate with twice the H frequency, but in the average the strength of K is self-regulating again to the slightly supercritical value of $\overline{K} \approx 1.2$. We may consider the transition to the limit cycle as a self-induced Hopf bifurcation where the value of c is self-regulating to the regime slightly above the bifurcation point.

The frequency of the limit cycle oscillation is modulated by the strength of the noise ξ^2. With varying noise strength the robot will execute an irregular searching behavior, i.e. the robot will move forward for some time then reverse velocity and move backwards and so on. The most interesting property however is observed when the robot collides with some obstacle so that the wheels get blocked. Then ξ^2 in eq. 8 is very large so that the rate of change of H largely increases and the robot nearly immediately will reverse its velocity. In this way the parameter dynamics may be said to create an explorative behavior which stays sensitive to (perturbations by) the environment.

In the applications with both wheels driven by homeokinetic neurons the robot is found to explore the world without getting stuck in corners or at other obstacles, see the videos [3]. Please note that these properties are not the performance of the trained neuron but instead result from the interplay of state and parameter dynamics, i.e. the concomitant effects of eqs. 1 and 8.

3 Several channels

Let us now consider the case of one controller neuron with the SM loop closed via several channels. The sensor values x_i may now depend in a more general form on the value of y. As the most basic model assumption we again use a simple proportionality so that we write the SM loop as

$$x_{i,t+1} = a_i y_t + \xi_i$$

where the deterministic part may be considered as the model and the "noise" ξ is the model error, the constants a_i being learned online. The update rule for the membrane potential is

$$\tau^{-1}\Delta z = -z + \sum_i c_i x_i + H \tag{9}$$

the FP of eq. 9 is at $z = Ky + H$ (we assume $\overline{\xi}_i = 0$ for all channels) where K

$$K = \sum_i c_i a_i$$

is the feed-back strength in the SM loop. The corresponding neuron output is at

$$y = \tanh\left(Ky + H\right) \tag{10}$$

The homeokinetic principle yields the following rules for the dynamics of the parameters

$$\begin{aligned} \Delta c_i &= \mu a_i - 2\mu z x_i - \gamma c_i \\ \Delta H &= -2\mu z \end{aligned} \tag{11}$$

where μ was introduced above, ξ^2 is the strength of the noise averaged over the channels, and γ (which is small) produces a (weak) decay of the weights. We again find that the change of c_i is given by a driving together with an anti-Hebbian term.

The parameter dynamics is analyzed in the following way. When starting with $c_i = 0$ for all i the feed-back strength K is zero so that in the beginning $y = 0$ and the driving term in the learning dynamics produces $\Delta c_i = \mu a_i$ hence $\Delta\left(c_i a_i\right) = \mu a_i^2$ so that the feed-back strength in each of the channels increases with channels of higher response strength $|a_i|$ being favored. Once the overall feed-back strength exceeds the critical value, the anti-Hebbian term comes into play and with $H = const$ (no update) convergence is reached if ever $1 = 2Ky^2$ under the constraint that the FP condition eq. 10 is satisfied. The stationary solution is reached if

$$\gamma c_i = \left(1 - 2Ky^2\right) a_i$$

or

$$K = \frac{1}{\gamma/a^2 + 2y^2} \tag{12}$$

so that $c_i = \alpha a_i$ with

$$\alpha = \frac{1}{\gamma + 2a^2 y^2}$$

K and hence α being obtained from the solution of the FP eq. 10 using eq. 12, i.e. from the solution of ($H = 0$ at present)

$$y = \tanh\left(\frac{y}{\gamma/a^2 + 2y^2} + H\right)$$

The position of the FP now depends on the value of γ/a^2. For instance the FP is at $y = 0.58$ and $K = 1.15$ if $\gamma/a^2 = 0.2$ which is only slightly lower as compared to $y = 0.65$ and $K = 1.21$ for $\gamma = 0$. Obviously each sensor i is integrated into the SM loop according to its response factor a_i. The inclusion of the H dynamics will again produce the limit cycle behavior but the result will stay valid in the average over (at least) one period. Note that the value of $\alpha > 0$ since $K > 0$.

Both this simple analysis and computer simulations readily show that the system is self-regulating again into the limit cycle oscillations independently on the number of channels and the values of a_i, since the feed-back strength is determined again by K alone. Hence we may say that the irregular, environment sensitive explorative behavior is reproduced also in the case of many channels where each channel is related to a single sensor. Moreover the c_i are found to reach values so that all sensors are integrated into the SM loop.

4 Switching sensors

One point of interest in the present paper is with the switching of sensors. In the above considerations we have assumed that the sensor response is essentially proportional to the velocity of the robot. This is not the case for a proximity sensor. However we could use a preprocessing and consider the change of the sensor value in the time step as one of our x_i. However the problem is that this sensor characteristics is valid only if the sensor is "on", i.e. the obstacle is in the reach of the sensor. Another point of interest of the present approach is the case that new sensors are installed for some time or that sensors temporarily break down. In any case our learning is to include the temporary switching on and off of sensors.

In principle the parameter dynamics of eq. 11 may well cope with the situation. Assume we switch on a new sensor k and assume the value of the coupling $c_k = 0$. Then in the beginning we have $\Delta(c_k a_k) = \varepsilon\xi^2 a_k^2 (1 - 2Ky^2) > 0$ since the damping term $-\gamma\varepsilon\xi^2 c_k a_k$ in this channel is negligible as compared to the other channels, because of the small value of c_k. Obviously the value of $c_k a_k$ is rising and this procedure will stop only if $c_k = \alpha a_k$ is reached. Concomitantly the couplings of the other sensors and hence the value of α is readjusted so that the global balance is reestablished. Hence a newly switched on sensor is automatically integrated into the SM loop. We may also see that the switching off is dealt with in the same automatic fashion.

The problem however is that the processes of readjusting the coupling vector c takes some time. In practical applications (see below) the switching on and off of sensors may take place in very short time intervals. It is therefore of interest not to relearn the couplings but instead to have a kind of long-term memory where the couplings are stored and are read out appropriately. This is possible either on the basis of direct information on the state of the sensors or on context information which is able of qualifying the sensor situation. We will study this case in the following.

5 Experiments with a physical Khepera robot

Our experiments have been chosen so that we can demonstrate both the effects described above, i.e. on the one hand we want to show that the robot adapts its exploration according to its knowledge of the world. This means that it will move more or less tentatively as long as the knowledge is small, i.e. the modeling error is large and with increasing knowledge a more and more explorative behavior

will originate. On the other hand we want to study the situation of switching sensors. We therefore consider a Khepera robot in a moveable box which on its hand is confined in a larger area with fixed walls as borders, see Figure 2.

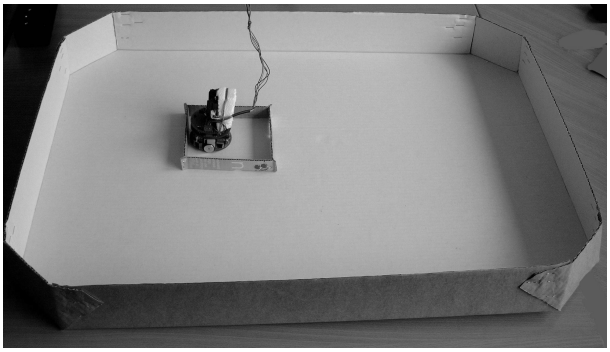

Figure 2. Khepera robot inside moveable box which again is in a larger box with fixed walls.

Two sensors are used, one measuring the velocity of the wheels $x_1(t) = v(t)$ and a pseudo-infrared (pseudo-IR) sensor. The output $x_2(t) = p(t)$ of the pseudo-IR sensor is proportional to the target velocity y of the robot so that it can be used immediately as one of the x_i in the parameter rule given above. However the pseudo sensor is triggered by the physical IR sensors with outputs $r(t)$ (the moveable box is large enough to contain both *on* and *off* regions of the pseudo IR)

$$p(t) = \begin{cases} by(t-1) & : \quad \max_{i=0}^{7} r_i(t) > r_{min} \\ 0 & : \quad \max_{i=0}^{7} r_i(t) \leq r_{min} \end{cases} \tag{13}$$

so that the sensor $p(t)$ is active as long as there is at least one IR sensor with an activity larger then the threshold r_{min}, else the value of this sensor will be zero. Thus we have the case of a switching sensor.

The controller of the robot consists of a single homeokinetic neuron with the update rule for the membrane potential as given above. Its output $y \in [-1, 1]$ is the forward velocity of the robot, steering is disabled. We introduce an auxiliary input $x_0 = 1$ so that the synapse c_0 corresponds to the threshold value H. So our vector of sensor values is $x = (1, v, p)^T$, the vector of response strengths is $a = (1, a_1, a_2)^T$ and the above formula runs over $i = 1, 2$ (including $i = 0$ in the case of the membrane potential).

We used several tricks in order to make the learning more effective. On the one hand the learning of c_0 is much faster then that of the other c since this one realizes the quick reaction of the robot to large model errors and hence the frequency effect. On the other hand with such large learning rates we must take some precautions against divergencies. Therefore we push each update through a squashing function, cf.

$$\Delta c \leftarrow \eta \tanh\left(\frac{1}{\eta}\Delta c\right)$$

so that there is no change in the case of small Δc but there is a maximum step width given by η. Moreover we use the derivative of the $\tanh$ function as $\tanh'() = 1.1 - \tanh^2()$ in order to have still non-vanishing updates even in the region of the saturation of the neuron.

Finally, the model parameters a_1 and a_2 are learned on-line by gradient descending the model error eq. 4. However in the model the measured speed depends linearly on the controller output, i.e. the model is appropriate only if the robot moves without problems. When colliding with the wall the model is not valid any longer and the learning should be switched off. This is achieved in our case by

multiplying the learning rate by a kind of reliance factor which is chosen as

$$r_j = \exp\left(-\beta \xi_j^2\right)$$

for channel j where $\xi_j^2 = (x_j - a_j y)^2$ is the error of the model in this channel.

5.1 Long-term memory for the parameters

A central point of the present work is that of switching sensors. In particular our pseudo-IR sensor $p(t)$ can switch frequently between $p\,(t) = 0$ and $p\,(t) = by(t)$ (see eq. 13) according to the robots position in the moveable box (Figure 2). So good predictions are obtained only if the model parameter is switched according to the situation.

The sensor situation depends on the output of the physical IR sensors r_i which can be transformed in a context $m(t)$ with the values

$$m(t) = \begin{cases} 1 & \max_{i=0}^{7} r_i(t) > r_{min} \\ 0 & \max_{i=0}^{7} r_i(t) \leq r_{min} \end{cases}$$

Our solution to this problem is to train a neural network (with the context $m(t)$ as input) to set the value of the model parameter $a_2(t)$ according to the context. The learning signals are directly given by gradient descending the model error.

Moreover different model parameters a_i produce different synaptic weights, see the discussion above. Therefore the controller weights are also represented by a neural network (with the context $m(t)$ as input), except for the bias weight $c_0(t)$ which is changing rapidly (compared with the other weights) all the time and therefore does not need to be memorized.

With the incorporation of this long-term memory the homeokinetic controller is able to handle switching sensors.

5.2 Results

In the experiments the model parameter for the wheel channel was set by hand so that only the model for the pseudo-IR channel had to be learned. The initial value of $a_2(t)$ was set to zero and the learning of the model parameter was disabled for the first 1000 steps. As a result the model error is large if the pseudo IR is active which leads to an almost immediate change of the value of c_0 so that the robot changes its direction of motion. The effect is that the robot avoids collisions with the walls of the moveable box it is enclosed in.

When model learning takes place after step 1000 the model parameter $a_2(t)$ becomes more and more adapted (jumping between $a_2 = 0$ and $a_2 = b$, cf. Figure 3 (c)) which decreases the model error. Hence when approaching the wall of the moveable box the relearning of c_0 (Fig. 4 (a)) does not take place and the robot starts moving the box around.

Eventually when the robot reaches the wall of the arena the wheel error is large so that the rapid relearning of c_0 and hence the velocity reversal takes place at this collision event. In this way the robot now explores the full region of the arena, see Fig.3 (a).

Approx. with step 4000 the learning is disabled and the model parameter $a_2(t)$ is set to zero again. This leads to a large model error and fast relearning of $c_0(t)$ when the pseudo-IR sensor is active. Hence the robot moves only in a short range of his environment like in the beginning of the experiment.

At the beginning and at the end of the experiment the pseudo-IR channel is not included in the SM loop ($c_2(t) \approx 0$, Fig. 4(b)) because $a_2(t)$ is set to zero. Hence the SM loop is closed only over the

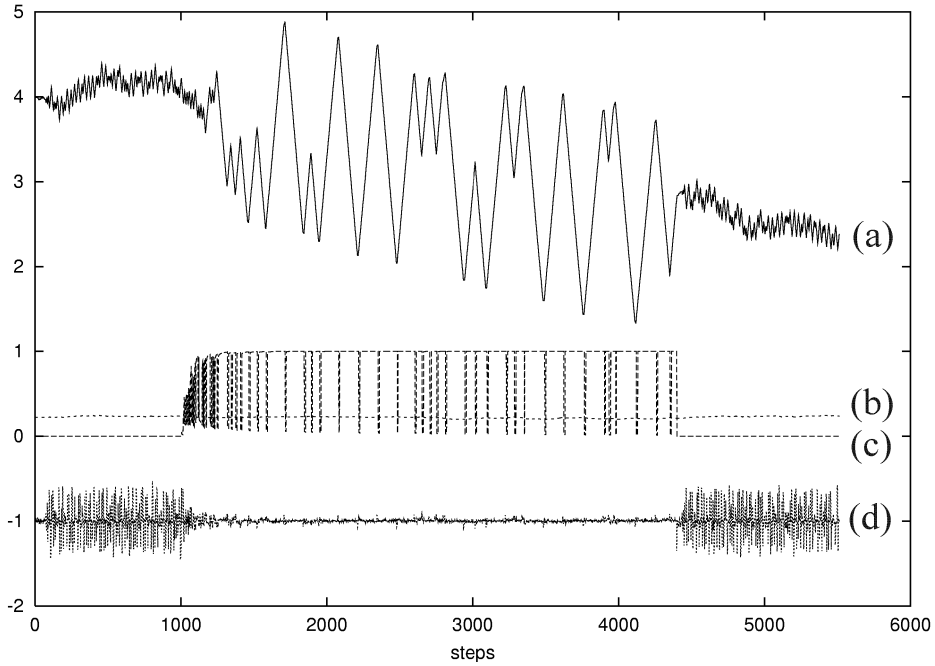

Figure 3. (a) The path traveled by the robot (with odometry error). For the first 1000 *steps model learning was disabled so that the robot moves only very cautiously not pushing the moveable box. Then model learning is allowed and with increasing knowledge of the world the robot starts moving the box and explores the full range of the arena while moving the box around. In the end learning was disabled (with* $a_2 = 0$*) again so that the cautious behavior reappears. (b) Model parameter* $a_1(t)$ *of the wheel channel is already learned at the beginning of the experiment. (c) Model parameter* $a_2(t)$ *unlearned at the beginning, then learning to jump between* 0 *and* 1 *depending on the activity of the IR sensors and reset to zero at the end. (d)* $\xi(t) - 1$*; The difference between predicted and measured pseudo-IR-sensor values (*$\xi(t)$*) gets smaller when the model learns to predict (middle) and rises with resetting the model parameter* $a_2(t)$ *to zero.*

wheels with $c_1(t) \approx 5$ (Figure 4 (c)) and with $a_1(t) \approx 0.23$ this leads to a feed-back strength $K \approx 1.1$ ($\alpha \approx 21.7$).

In the middle part the pseudo-IR sensor is included in the SM loop, but only when it is active. Then the model parameter $a_2(t) \approx 1$ leads to the increase of the appropriate weight until $c_2(t) \approx 1$. The readjusted factor $\alpha \approx 1$ can also be seen in the wheel channel ($c_1(t) \approx 0.2$, $a_1(t) \approx 0.23$) so that K is around 1.1 again. When the pseudo-IR sensor is not active the parameters should be $a_2(t) \approx c_2(t) \approx 0$ and $c_1(t) \approx 5$. This is not the case ($c_2(t) \approx 1$, $c_1(t) \approx 3.3$) so the feed-back strength is smaller then 1 for a short time. The cause is seen in the very fast switching of the pseudo sensor. The membrane potential z is large when entering the region with pseudo IR off, because the robot is moving across the moveable box. With the slow changing of z (cf. eq. 9) the neuron output y remains large. Hence the anti-Hebbian term $2Kx_iy = 2Ka_iy^2$ in eq. 11 stays more dominant for the short time the pseudo sensor is off leading to a smaller value of c_1 then expected.

In a few words this experiment shows the uncertain, tentative behavior (changing direction of motion very often) of the robot in areas the model can not predict properly, the ”brave” behavior (covering large areas, which are predictable for the model) and the change between these two behaviors through learning of the model parameters. And by the way the switching sensor is integrated in the SM loop, as long as it is active.

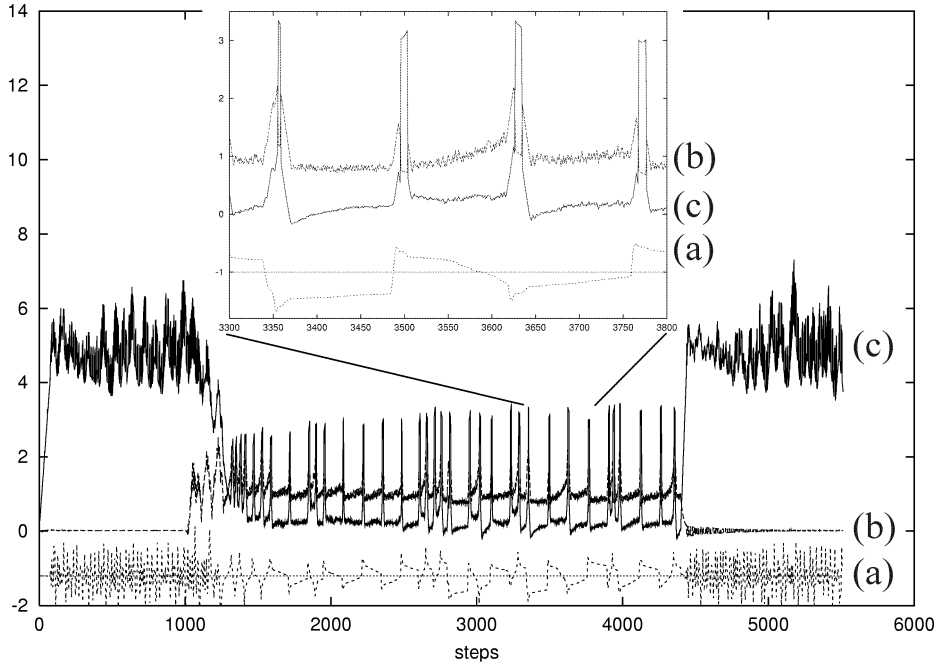

Figure 4. (a) The bias weight $c_0(t) - 1$, (b)the weights for the pseudo-IR input $c_2(t)$ and (c) the wheel input $c_1(t)$ of the homeokinetic neuron used as controller. The input weights converge to different values depending on the value of the model parameter for the pseudo-IR channel. The frequency of the changes of $c_0(t)$ and therefore the time the robot travels in one direction depends on the model error.

6 Outlook and Conclusions

We have demonstrated in the present paper that the very simple learning rules, cf. eq. 8 derived from the homeokinetic principle realize a self-learning robotic system which can survive in a sufficiently simple world without any further external help. In particular we have seen that sensors are automatically integrated according to their response strength as soon as they deliver a signal to the controller. Moreover the system also can deal with the problem of a rapid change in the properties of the sensors. The basic effect observed is that the learning rule drives the robot into an explorative mode of behavior which however is sensitive to the reactions of the environment by way of the model error. From a dynamic systems point of view we have a closed loop control system with a pitchfork or a Hopf bifurcation (if the learning of the threshold is included) and the effect of the learning is to drive the system to a regime slightly above the bifurcation point where such systems are known to be particularly sensitive.

The findings of the present paper reveal also a general property of our approach which is related to the dilemma between exploration and the exploitation of knowledge gathered in the model. Typically this difficulty arises in the following way. Assume we have an agent which is to explore the world and while exploring to construct a model of its behavior which on its hand is used for the guidance of control. Then one can either choose to stay with the behaviors which are well modelled and be safe or to further explore the space with the chance of getting lost. This needs a careful tuning of the exploration rate in practical realizations.

In our approach this dilemma is solved by the fact that the robot reacts with a change in behavior if the model error is increasing suddenly. This may happen if the robot gets into a situation which is

not yet ”understood” by its model. This restricts the robot to the behavior which is well modelled by the current model. However with each of such situations the model also gets some new information which can be integrated into the model making the error smaller and hence the reaction slower when the robot reencounters the situation. In this way the robot will conquer increasingly larger regions of the behavior space. In other words the homeokinetic controller automatically adapts the exploration rate to the needs of (model) knowledge acquisition.

Another aspect is in the relation between the short term memory of the original setting where the parameters of the neuron are changed on the behavioral time scale and the long-term memory introduced in the present paper. We have demonstrated that the rapid parameter learning can well be used as a learning signal for a long-term memory which stores the parameter values of the neuron in a context dependent fashion so that a parameter and hence behavior recall guided by context information is realized. In future work this technique will be used for a weighted feed-in of the learning signal into the parameter network the weights being given by a kind of reinforcement signal obtained either from outside or in terms of the future model errors. In the latter case on might use the wheel error so that behaviors which lead to collisions with obstacles would be avoided as a result of the learning. Hence we may expect that the agent acquires more foresight in the course of the time.

References

[1] W. B. Cannon. *The Wisdom of the Body*. Norton, New York, 1939.

[2] R. Der. Self-organized acquisition of situated behavior. *Theory Biosci.*, 120:179–187, 2001.

[3] R. Der. Videos on homeokinetic control of robots. http://www.informatik.uni-leipzig.de/˜der/Forschung/videos.html, 2003.

[4] R. Der, M. Herrmann, and M. Molicki. Self-organization in sensor-motor loops by the homeokinetic principle. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, page 510, 1 2002.

[5] R. Der, U. Steinmetz, and F. Pasemann. 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, 1999. IOS Press.

[6] R. Pfeiffer and C. Scheier. *Understanding Intelligence*. MIT Press,Cambridge, MA, 1999.

[7] W. R.Ashby. *Design for a Brain*. Chapman and Hill, London, 1954.