Chapter 2
Learning Legged Locomotion

Fumiya Iida and Simon Bovet

2.1 Introduction

Legged locomotion of biological systems can be viewed as a self-organizing process of highly complex system–environment interactions. Walking behavior is, for example, generated from the interactions between many mechanical components (e.g., physical interactions between feet and ground, skeletons and muscle-tendon systems), and distributed informational processes (e.g., sensory information processing, sensory-motor control in central nervous system, and reflexes) [21]. An interesting aspect of legged locomotion study lies in the fact that there are multiple levels of self-organization processes (at the levels of mechanical dynamics, sensory-motor control, and learning).

Previously, the self-organization of mechanical dynamics was nicely demonstrated by the so-called Passive Dynamic Walkers (PDWs; [18]). The PDW is a purely mechanical structure consisting of body, thigh, and shank limbs that are connected by passive joints. When placed on a shallow slope, it exhibits natural bipedal walking dynamics by converting potential to kinetic energy without any actuation. An important contribution of these case studies is that, if designed properly, mechanical dynamics can generate a relatively complex locomotion dynamics, on the one hand, and the mechanical dynamics induces self-stability against small disturbances without any explicit control of motors, on the other. The basic principle of the mechanical self-stability appears to be fairly general that there are several different physics models that exhibit similar characteristics in different kinds of behaviors (e.g., hopping, running, and swimming; [2, 4, 9, 16, 19]), and a number of robotic platforms have been developed based on them [1, 8, 13, 22].

Dynamic interactions of distributed information processing also play an important role in stable and robust legged locomotion, which has previously been shown in the locomotion studies of biologically inspired motor control architectures, the so-called central pattern generator models (CPGs; [14]). This approach typically

A. Adamatzky and M. Komosinski (eds.), *Artificial Life Models in Hardware*,

simulates the dynamic interactions of neurons, and the periodic oscillatory signal output of the neural network is connected to the motors of legged robots. Because of the dynamic stability in the signal output, the locomotion processes using this architecture generally exhibit robust locomotion of complex musculo-skeletal structures [20, 25], and it has been shown that the legged robots are capable of legged locomotion in relatively complex environment [7, 10, 15, 17].

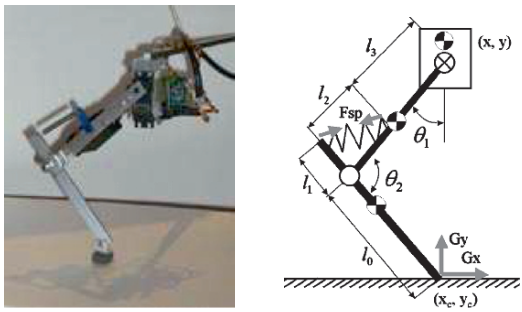
As exemplified in these case studies, one of the most challenging problems in the studies of legged locomotion is to identify the underlying mechanisms of self-organization that induces physically meaningful behavior patterns in terms of stability, energy efficiency, and controllability, for example, [3]. From this perspective, the goal of this article is to explore how the self-organization processes in the physical system–environment interactions can be scaled up to more "intellignet" behaviors such as goal-directed locomotion by discussing two case studies of learning legged robots. More specifically, while the dynamic legged locomotion research were limited to only periodic behavior patterns, we will explore the mechanisms in which the rules of motor control can be generated from the physical interactions in the legged robotic systems. Note that this article shows only the important aspect of the case studies in order to discuss conceptual issues. More technical details can be found in the corresponding publications [6, 11].

2.2 Learning from Delayed Reward

Physical dynamic interactions play an important role not only for the repetitive behavior patterns such as walking and running on a flat terrain, but also for the resilient behaviors such as high jumps and kicking a ball. Generating such resilient behaviors generally involves nonlinear control that requires a certain form of planning. For example, a high jump requires a preparation phase of several preceding steps; ball-kicking requires a swing back of the leg in a specific way to gain the maximum momentum at impact. The optimization of such behavior control can be characterized as a "delayed reward" learning problem [24], which means, for example, that a system can realize it was a bad step only after falling over. To deal with such nonlinear control of body dynamics, this section explores a case study of a one-legged hopping robot that learns to generate a series of high-jumps to traverse a rough terrain [11].

2.2.1 One-legged Hopping Robot

Figure 2.1 shows one of the simplest legged robot models. This robot consists of one motor at the hip joint and two limb segments connected through an elastic passive joint. This system requires only a simple motor oscillation to stabilize itself into



(a) (b)

Fig. 2.1 (**a**) Photograph and (**b**) schematic of the one-legged hopping robot. It consists of one servomotor at the hip joint (represented by a *circle with a cross*) and two limb segments connected through a compliant passive joint (marked by an *open circle*)

a periodic hopping behavior [23]. The hip motor uses a position feedback control, in which the angle of hip joint is determined by three parameters: amplitude A, frequency f, and offset of oscillation B.

$$P(t) = A\,\sin(2\pi f t) + B. \tag{2.1}$$

When these parameters are set properly, the robot shows stable periodic hopping behaviors (Fig. 2.2), and behavioral characteristics resulting from its particular morphology can be summarized as follows. First, locomotion can only be achieved dynamically. As the leg has one single actuated degree of freedom, the only way the robot can lift its legs off the ground is by delivering enough energy through the motors to make the whole body jump. Second, stability is achieved through the material properties of the legs (especially the compliance of the passive joints) rather than by actively controlling the positions of all joints. For instance, an inadequate position of the lower limb (which is only passively attached to the upper limb) during the flight phase will automatically be corrected by the spring on contact with the ground. In particular, this characteristic allows the robot to be controlled in an open-loop manner (i.e., without any sensory feedback) over a continuous range of control parameters. By simply actuating periodically the motors back and forth, the robot put on the ground will automatically settle after a few steps into a natural and stable running rhythm. Third, the elasticity of the legs, partially storing and releasing energy during contact with the ground, allows to achieve not only stable, but also rapid and energy efficient locomotion. The importance of such elastic properties in muscle–tendon systems has been long recognized in biomechanics, where it has a particular significance in theoretical models for the locomotion of legged animals [2, 19].

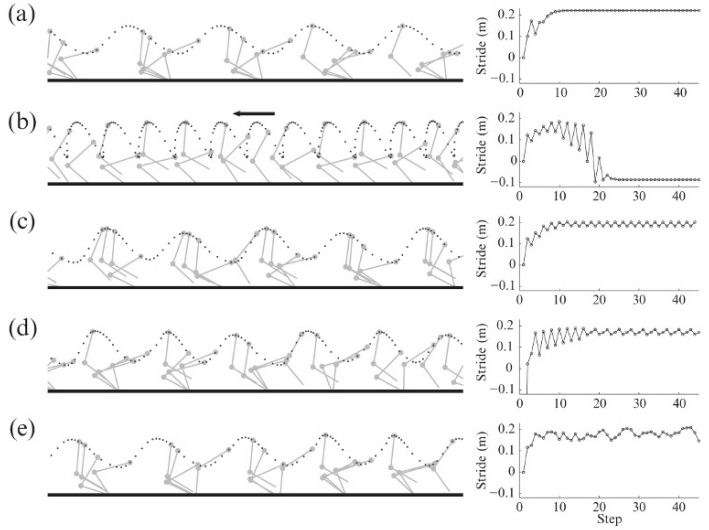


Fig. 2.2 Self-stability and variations of locomotion processes: (**a**) stable forward locomotion with a constant stride length, (**b**) backward locomotion, (**c**) and (**d**) forward locomotion with two and three step cycles, (**e**) stable locomotion with chaotic stride lengths. The oscillation frequencies of the hip motor are $f = 2.78$, 2.72, 2.85, 2.78, and 2.73 Hz (from top to bottom)

2.2.2 Learning to Hop Over Rough Terrain

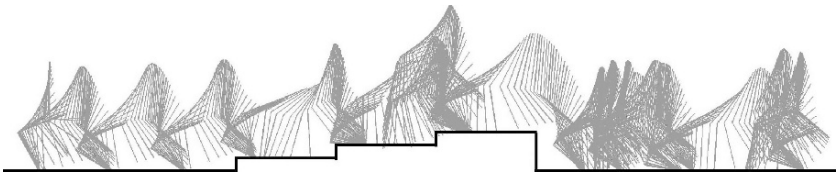
Although the periodic dynamic locomotion can be mechanically stabilized against small disturbances, the hopping robot needs to actively manipulate the motor control parameters to deal with more complex environment. In this experiment, we applied a machine learning method, the so-called Q-learning algorithm [24], for optimizing the oscillation frequency of the actuated joint.

The learning process repeats locomotion experiments in a given environment until it reaches to a certain number of leg steps. In a learning step n, the system tests a motor control policy consisting of a series of motor oscillation frequencies $f_{i=1,2,\ldots}$, which is determined from a probability matrix $Q^n(i,f)$ (i is the number of leg steps). After each trial, the learning process receives a positive reward signal proportional to the traveling distance and negative reward in case the robot falls over.

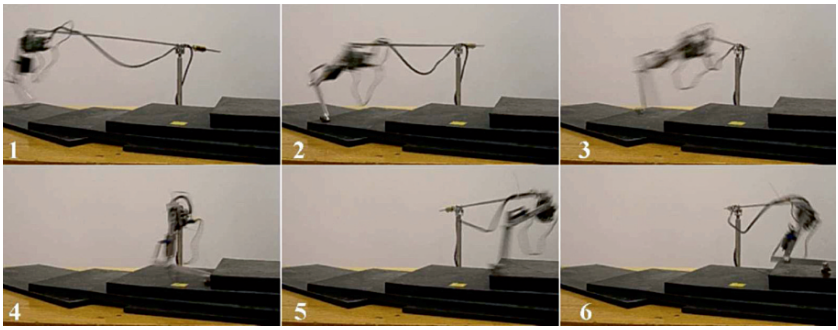
$$R^n(i) = \begin{cases} -5.0 & : i = \text{FailedStep} \\ \text{FinalDistance} & : i \neq \text{FailedStep} \end{cases} \tag{2.2}$$

The learning process then updates the probability matrix with a certain learning rate α as follows:

$$\begin{aligned} Q^{n+1}(i,f_i) = (1.0-\alpha)Q^n(i,f_i) + \alpha(R^n(i) \\ +\gamma \max(Q^n(i+1,f))) \end{aligned} \tag{2.3}$$



(a)



(b)

Fig. 2.3 (**a**) Learning results of motor control in simulation. The optimized sequence of motor frequencies exhibits 12 leg steps successfully traveling through a rough terrain. (**b**) Time-series photographs of the robot hopping over the steps. The motor control parameter was first optimized in simulation and transferred to the robot for the real-world experiment

As in the typical reinforcement learning, this learning process utilizes a discount factor γ, which influences the selection of action with respect to the prior action. For example, when the action f_i at a leg step i resulted in a successful continuation of locomotion, the learning process reinforces the probability of choosing f_{i-1} with a discount factor γ as well as that of f_i.

The hopping robot was implemented in a physically realistic simulator to facilitate a number of trials and errors in the learning process, and the learned parameters were transferred to the real-world robotic platform. After a few hundred iterations in the learning phase, the system is able to find a sequence of frequency parameters that generates a hopping gait of several leg steps for the locomotion of the given rough terrain (Fig. 2.3).

Searching for a specific series of frequency parameters is not a trivial problem, because the choice of parameter not only influences behavior of the corresponding leg step, but also those of subsequent leg steps. For example, if the system changes a control parameter at the leg step i, the exactly same motor output of the leg step $i+1$ often results in completely different behaviors. It is, therefore, necessary to utilize the delayed-reward learning such as the Q-learning algorithm explained above, and the typical characteristics in the learning process is illustrated in Fig. 2.4. At the earlier learning steps, the robot attempts mostly random sequences of the control parameters, which are more structured at the later stage. The search process is, however, not straight forward in a way that, at a certain learning step, the control parameters at earlier leg step is modified to achieve breakthroughs. For example,

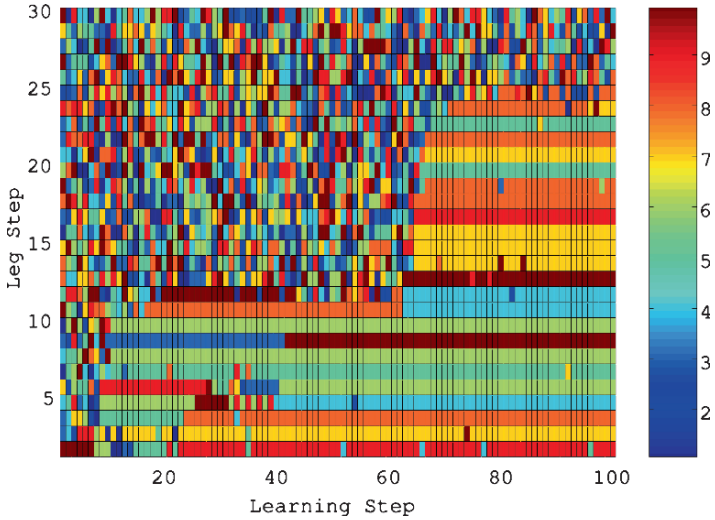


Fig. 2.4 A learning process of motor control policies. The color in each tile indicates the oscillation frequency of motor at the leg step N. It is shown that the control policy is structured towards the end of the learning process

while the learning process in Fig. 2.4 could not find the adequate parameter at the leg step 12 (after the learning step 17), it had to explore the parameter space until the parameter change at the leg step 9 (at the learning step 42), which eventually resulted in a breakthrough to continue the locomotion thereafter.

In summary, this case study explored a learning architecture that exploits dynamics of a compliant leg for goal-directed locomotion in rough terrain. To achieve highly dynamic locomotion such as a series of high jumps over large steps, the learning architecture requires a self-organization process that explores time-series motor output: because behavior of the robot is not only dependent on an immediate motor output but also on the prior ones, the delayed-reward mechanism (the propagation of reward signals over multiple leg steps) is necessary in the learning architecture. It is important to note that the goal-directed behavior shown in this case study was a result of the two levels of self-organization processes (i.e., in mechanical and informational dynamics): because the learning process exploited the underlying mechanical self-stability, the basic forward locomotion dynamics do not require parameter optimization, on the one hand, and the rich behavioral diversity of various hopping heights can be generated only by manipulating frequency parameters.

2.3 Learning from Implicit Reward

The previous case study employed a rather simple setup of learning experiments to emphasize the roles of delayed-reward signals in a learning process of legged locomotion. In contrast, this section discusses how the complexity of self-organization

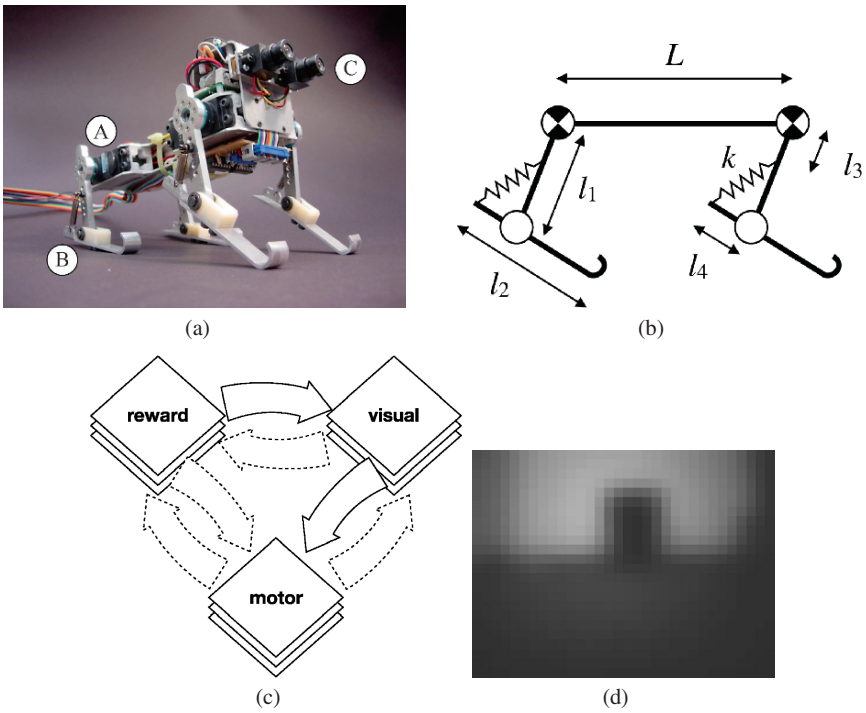


Fig. 2.5 (**a**) Four-legged running robot with two cameras (only one of them was used in the experiments). (**b**) Schematic of the robot, illustrating the locations of motors (*circles with crosses*) and passive joints (*open circles*). (**c**) Neural network architecture that illustrates synaptic weights. *Solid arrows* represents nonzero synaptic connections after the learning phase through which neural activities can propagate, while *dotted arrows* represent synaptic connections with essentially zero weights. (**d**) Snapshot image correlated to the reward signals

processes can be scaled up such that nontrivial signal pathways can be developed between sensory input and motor output. Here, we introduce another learning architecture that extracts correlation between signals to propagate implicit reward signals for a visually mediated target following behavior.

2.3.1 Four-legged Running Robot

The robotic platform used in this case study is a running robotic dog [12] shown in Fig. 2.5. This robot has four identical legs, each of which consists of one servomotor actuating a series of two limbs connected through a passive elastic joint as in the previous case study. Dynamic locomotion is also achieved by periodically moving back and forth the servomotors actuating the legs of the robot, and the target angular position $P_i(t)$ of motor i at time t is given by

$$P_i(t) = A_i \sin(2\pi f t + \phi_i) + B_i, \tag{2.4}$$

where A_i is the amplitude and B_i the set point of the oscillation. Compared with the one-legged hopping robot, this robot has a few additional parameters ϕ_i, the phase offsets, which determine the phase delay of oscillation between the legs.

The learning architecture of this robot has a form of neural network, which receives signals online from a visual sensor and provides output signals to the control parameters of (2.4). The neural network is specifically designed for extracting correlations in sensory-motor signals by using a modified Hebbian learning rule (see [6] for more details). We have modeled three groups of neurons, that is, motor neurons, sensor neurons, and "reward neurons," which are fully connected internally as shown in Fig. 2.5c.

The motor neurons are connected to a set of motor variables that represent the *differences* of parameter values between left-side and right-side motors, as well as between fore and hind motors. For instance, the oscillation amplitudes A_i of the four motors are defined as follow:

$$A_{\text{fore,left}} = A_0 - \frac{1}{2}\Delta A_{\text{lat}} - \frac{1}{2}\Delta A_{\text{long}} \tag{2.5}$$

$$A_{\text{fore,right}} = A_0 + \frac{1}{2}\Delta A_{\text{lat}} - \frac{1}{2}\Delta A_{\text{long}} \tag{2.6}$$

$$A_{\text{hind,left}} = A_0 - \frac{1}{2}\Delta A_{\text{lat}} + \frac{1}{2}\Delta A_{\text{long}} \tag{2.7}$$

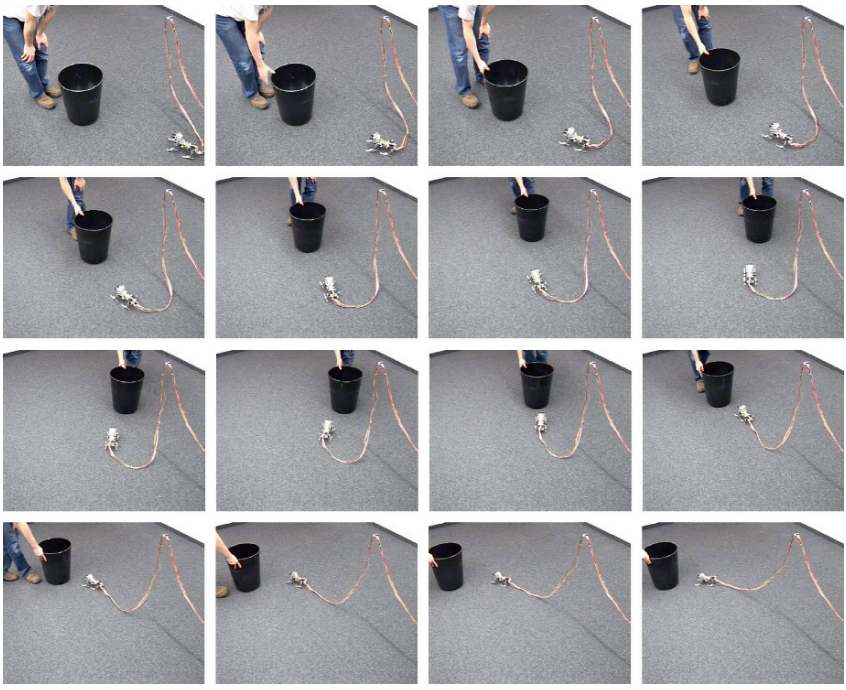
$$A_{\text{hind,right}} = A_0 + \frac{1}{2}\Delta A_{\text{lat}} + \frac{1}{2}\Delta A_{\text{long}} \tag{2.8}$$

ΔA_{lat} and ΔA_{long} are the lateral and longitudinal differences of amplitude, and A_0 is the average amplitude. The other motor parameters (i.e., the set points B_i and the phase offsets ϕ_i) are defined accordingly. Eventually, we have the eight state components (i.e., A_0, ΔA_{lat}, ΔA_{long}, B_0, ΔB_{lat}, ΔB_{long}, $\Delta\phi_{\text{lat}}$, and $\Delta\phi_{\text{long}}$) whose values are represented by the activity of eight motor neurons. Note that the frequency of oscillation f is constant for all motors, which provides a basic setup of the robot running forward.

The robot is equipped with a vision system consisting of a camera attached to the body and pointing in the forward direction (see Fig. 2.5). The sensor neurons are receiving both intensity and estimated optical flow extracted from the gray-scale visual input of the 32×24 pixel values. For enabling reinforcement learning, we also include a set of "reward neurons" as described later.

2.3.2 Learning to Follow an Object

The experiment of this case study consists of two phases. In the initial phase, the motor neurons are randomly activated, thus producing arbitrary motions of the robot. This initial phase allows the neural network to learn the basic cross-modal



(a)

Fig. 2.6 Four-legged running robot following a black object in an unstructured environment

correlations as follows. The reward is delivered when the robot is facing a large black bin placed in the environment (as shown in Fig. 2.6). The synaptic connections between the reward and the sensor neurons therefore learn a correlation between reward signals and a set of visual input signals that correspond to the image of the black bin in the center of the visual field (as shown in Fig. 2.5d). At the same time, because of the Hebbian-like learning rule, the synaptic connections between the visual and motor groups of neurons capture another significant correlation. This correlation, which we will elaborate later, involves the visual neurons that receive optical flow signals and a particular pattern of activity in the motor neurons. In the second phase, the robot is let to move on its own while activating the sensory neurons receiving reward signals. Because of the particular synaptic connections that have been strengthened during the initial phase of the experiment, the reward signals are propagating through visual neurons to motor neurons, which eventually activate the oscillation of the legs such that the robot follows the object.

The observed behavior, generated from the propagation of neural activity across the network, is illustrated in Fig. 2.6, where the robot turns towards any black object that is placed in the center of its field of view and follows the object as it is moved around. The key aspect of the network connectivity is the correlation between perceived visual flow and motor activity, which is captured by the synaptic

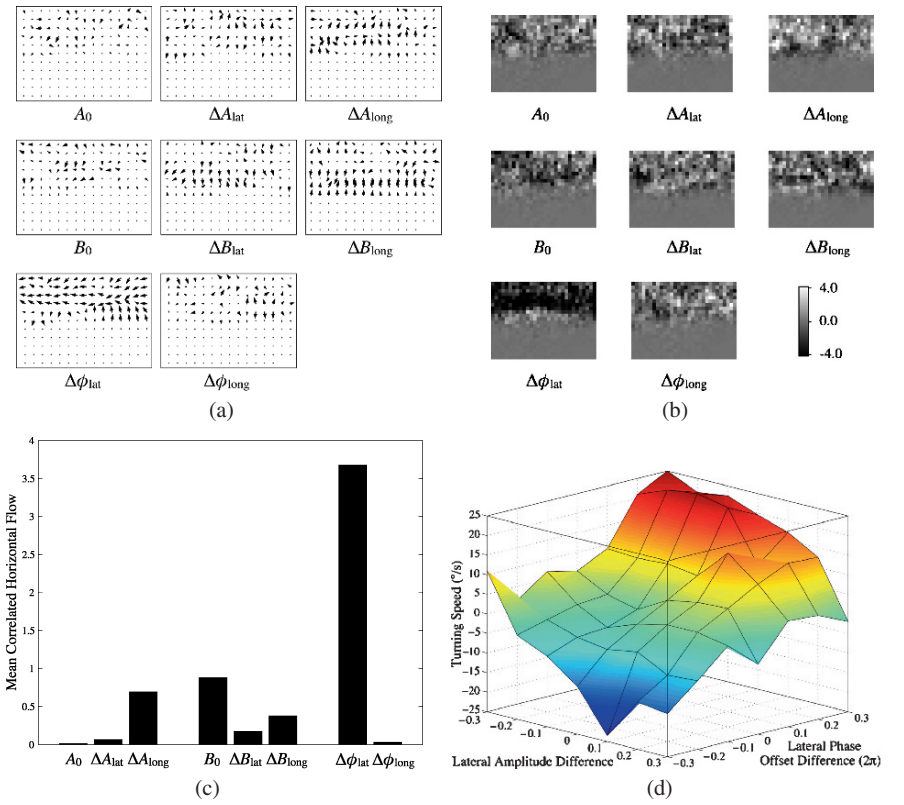


Fig. 2.7 Graphical representation of the synaptic weights coupling the visual modality to the motor modality, showing (**a**) the visual flow field, and (**b**) only the horizontal component thereof, correlated to each of the eight components of the motor state. (**c**) Average horizontal component of the visual flow correlated to each motor component (absolute value). (**d**) Turning speed of the robot as a function of both lateral amplitude difference ΔA_{lat} and later phase offset difference $\Delta\phi_{\text{lat}}$

weights coupling the visual modality to the motor modality. Figure 2.7 shows a graphical representation of these weights, illustrating the visual flow correlated to each motor control parameter. Clearly, the neural architecture captures a significant correlation only between visual flow and the motor parameter corresponding to lateral phase offset difference ($\Delta\phi_{lat}$). This means that the quadruped robot learns a control strategy for turning that modifies essentially the phase difference between the oscillations of the left and the right legs.

To better understand this result, we systematically quantify the turning rate of the robot as a function of various motor control parameters. Figure 2.7d shows that the turning speed is most easily and robustly controlled with the lateral phase difference, the relation between the two quantities being almost linear in the considered range. In contrast, when the other motor control parameters are varied, the turning speed of the robot either does not change significantly or displays no linear relation: for

instance, as the lateral amplitude difference is steadily increased, the robot does not always change the turning rate monotonously.

In summary, the modified Hebbian learning rule, which captured the correlation patterns of sensory-motor activity in the neural network, developed a nontrivial synaptic structure that produces an object following behavior based on the visual sensory input. To achieve this task, the network has to find a nontrivial correlation between visual sensory input, reward signals, and motor signals. This experiment shows how self-organization processes that capture correlation of sensory-motor signals can generate sensible behavior patterns.

2.4 Conclusion

This article discussed the issues of legged locomotion from the perspective of artificial life in the real world. By treating legged locomotion as a self-organizing process resulting from complex physical and informational dynamics, we argue that one of the most significant challenges lies in the grounding of self-organization processes for physically meaningful behaviors. While our exploration is still at a nascent stage, we extracted a few important principles from the case studies presented in this article. In particular, we have shown that a learning architecture requires, on the one hand, reward signals evaluating a series of motor actions to make full use of nonlinear mechanical dynamics, and on the other, a specific form of signal propagation to capture the patterns of sensible physical system–environment interactions for goal-directed behaviors.

There are still a number of open questions that we have not explicitly discussed in this article so far. One of the fundamental questions is how we could extend the complexity of self-organizing processes further with less "hand-coded" elements in the embodied systems. For example, in the case studies presented in this article, we predefined a number of elements such as the basic controllers that generate sinusoidal oscillation, basic sensory information processing (e.g., optical flow estimation), mechanical dynamics with fixed viscous-elasticity in passive joints, and the basic reward signals, to mention but a few. Although we found these predefined elements essential to maintain the learning phase within a reasonable amount of time, it requires further studies to discuss how the self-organizing processes should be structured. We are expecting that the comparative study with some of the related work (e.g., [5, 17, 26]) will clarify more general rules to manage the higher dimension of parameter space in self-organizing processes of embodied systems.

Acknowledgment

This work is supported by the Swiss National Science Foundation (Grant No. 200021-109210/1) and the Swiss National Science Foundation Fellowship for Prospective Researchers (Grant No. PBZH2-114461).

References

1. Ahmadi, M., Buehler, M.: Controlled passive dynamic running experiments with ARL monopod II. IEEE Transactions on Robotics, 22, 974–986 (2006)
2. Alexander, R.McN.: Three uses for springs in legged locomotion. International Journal of Robotics Research, 9, 53–61 (1990)
3. Bedau, M.A., McCaskill, J.S., Packard, N.H., Rasmussen, S., Adami, C., Green, D.G., Ikegami, T., Kaneko, K., Ray, T.S.: Open problems in artificial life. Artificial Life 6, 363–376 (2000)
4. Blickhan, R., Seyfarth, A., Geyer, H., Grimmer, S., Wagner, H.: Intelligence by mechanics. Philosphical Transactions of the Royal Society of London Series A: Mathematical and Physical Sciences 365, 199–220 (2007)
5. Bongard, J., Zykov, V., Lipson, H.: Resilient machines through continuous self-modeling. Science 314, 1118–1121 (2006)
6. Bovet, S.: Robots with self-developing brains. Dissertation, University of Zurich (2007)
7. Buchli, J., Ijspeert, A.J.: Self-organized adaptive legged locomotion in a compliant quadruped robot. Autonomous Robots 25, 331–347 (2008)
8. Collins, S., Ruina, A., Tedrake, R., Wisse, M.: Efficient bipedal robots based on passive dynamic walkers. Science 307, 1082–1085 (2005)
9. Dickinson, M.H., Farley, C.T., Full, R.J., Koehl, M.A.R., Kram, R., Lehman, S.: How animals move: An integrative view. Science 288, 100–106 (2000)
10. Geng, T., Porr, B., Wörgötter, F.: A reflexive neural network for dynamic biped walking control. Neural Computation 18, 1156–1196 (2006)
11. Iida, F., Tedrake, R.: Optimization of motor control in underactuated one-legged locomotion. International Conference on Robotics and Systems (IROS 07), 2230–2235 (2007)
12. Iida, F., Gomez, G.J., Pfeifer, R.: Exploiting body dynamics for controlling a running quadruped robot. Proceedings of International Conference on Advanced Robotics (ICAR 2005), 229–235 (2005)
13. Iida, F., Rummel, J., Seyfarth, A.: Bipedal walking and running with spring-like biarticular muscles. Journal of Biomechanics 41, 656–667 (2008)
14. Ijspeert, A.J.: Central pattern generators for locomotion control in animals and robots: A review. Neural Networks 21, 642–653 (2008)
15. Kimura, H., Fukuoka, Y., Cohen, A.-H.: Biologically inspired adaptive walking of a quadruped robot. Philosophical Transactions of the Royal Society of London Series A: Mathematical and Physical Sciences 365, 153–170 (2007)
16. Kubow, T.M., Full, R.J.: The role of the mechanical system in control: A hypothesis of self-stabilization in hexapedal runners. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 354, 849–861 (1999)
17. Matsubara, T., Morimoto, J., Nakanishi, J., Sato, M., Doya, K.: Learning CPG-based biped locomotion with a policy gradient method. Proceedings of 2005 5th IEEE-RAS International Conference on Humanoid Robots, 208–213 (2005)
18. McGeer, T.: Passive dynamic walking. The International Journal of Robotics Research 9, 62–82 (1990)
19. McMahon, T.A.: Muscles reflexes and locomotion. Princeton University Press, Princeton, NJ (1984)
20. Ogihara, N., Yamazaki, N. Generation of human bipedal locomotion by a bio-mimetic neuro-musculo-skeletal model. Biological Cybernetics 84, 1–11 (2001)
21. Pfeifer, R., Lungarella, M., Iida, F.: Self-organization, embodiment, and biologically inspired robotics. Science 318, 1088–1093 (2007)
22. Raibert, H.M.: Legged robots that balance. MIT Press, Cambridge, MA (1986)
23. Rummel J., Seyfarth A.: Stable running with segmented legs. International Journal of Robotics Research 27, 919–934 (2008)
24. Sutton, R., Barto, A.: Reinforcement learning. MIT Press, Cambridge, MA (2000)

25. Taga, G., Yamaguchi, Y., Shimizu, H.: Self-organized control of bipedal locomotion by neural oscillators in unpredictable environment. Biological Cybernetics 65, 147–159 (1991)
26. Tedrake, R., Zhang, T.W., Fong, M., Seung, H.S.: Actuating a simple 3D passive dynamic walker. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2004), 4656–4661 (2004)