Tacit Learning – Machine Learning Paradigm Based on the Principles of Biological Learning

Shingo Shimoda

Autonomous Behavioral Control unit, RIKEN, BSI-TOYOTA collaboration center, 2271-130,
Anagahora, Shimo-shidami, Moriyama-ku Nagoya, Aichi, 463-0003, Japan
shimoda@brain.riken.jp

1 Introduction

Adaptations to unpredictable environmental changes enable living organisms to survive in their natural environments and are therefore the highest-priority tasks for all of them. In the long history of evolution, living organisms have developed regulatory systems that can adapt their activities to the environment and, as a result have been able to extend their activity fields to almost all places on the earth.

One of the most curious features of biological regulatory systems is the method of computations for the regulations. All the computations in biological systems are carried out by the activities of homogeneous computational media, and networks of these computational media deal with all environmental inputs relevant to an organism's survival. Neurons in a brain[1][2], protein-protein interactions in intracellular regulations[3], and T- and B- cell activities in adaptive immune systems[4] are prominent examples of computational media.

The most important feature of these regulatory systems is that the accumulation of the local activities of the computational media can adapt the global activity of their network to the environment without any supervising signals from the global point of view. Let us take behavior control by the brain as an example. The brain is a network of neurons whose activity rules are genetically defined. The neurons modify their synaptic connections, the amount of neurotransmitters they release, the conductances of their ion channels, and so on according to their innate rules. Behaviors are gradually changed to ones more adapted to the environment as a result of the accumulation of each neuron's activities even though nothing is supervising the behavior adaptation from the global point of view.

Another important feature of these biological regulation systems is that the control and the adaptation progress without an explicit distinction between them[5]. The environmental information taken into the controller through the behavior control is used to tune the behaviors. This feature of biological regulatory systems makes it possible to adapt behaviors to the continuously changing natural environment in everyday life. Figure 1 is a conceptual image of the regulation-by-brain system. When a living organism moves, a feedback loop including the environment, the body, and the brain is constructed. The inputs from

S. Mohammed et al. (eds.), *Intelligent Assistive Robots*,
Springer Tracts in Advanced Robotics 106, DOI: 10.1007/978-3-319-12922-8_8

the environment stimulate the neurons in the brain through the sensors, and the activities of the neurons create behaviors that influence the environment. In this feedback loop, behaviors are gradually adapted to the environment in parallel with behavior creations by accumulating the activities of the neurons even though each neuron does not recognize the state of the whole environment.

Although recent studies have succeeded in clarifying the details of biological computational media[6]–[8] and their network structures[9][10], the process of organizing the activities of computational media and creating global behaviors adapted to the environment remains an attractive research topic. Some attempts have been made to associate the activities of computational media with the global functions of their networks[11]–[17]. These analyses, however, have clarified only a few functions of biological regulations in static environments or in environments with controlled changes.

The construction of artificial controllers that are capable of adapting to unpredictable changes is an interesting approach useful for understanding the adaptation process because adaptation processes in artificial systems can be easily analyzed than those in biological systems[18]. Recent advances in artificial learning and adaptive methods for robot control, however, have not reached the level of adaptability of biological systems[19]–[32]. One of the most critical problems with conventional approaches is the way of specifying goals of learning. The goals of learning in many cases are specified in advance by using supervising signals such as teaching signals in neural networks[19]–[21], cost functions in genetic algorithm[25][26], and reference signals in adaptive control[27][28], that are not changed during learning even though the environments have changed. It would be difficult to adapt to environmental changes in learning that was supervised by fixed signals. Nowadays, these is a rapidly growing social demand for robots as real partners of human beings in role such as caring for the elderly, disaster rescue and rehabilitation assistance, which obviously require high levels of adaptability. The development of artificial controllers with high levels of adaptability is now an imminent problem facing the field of robotics, not just for understanding biological systems.

In this chapter we discuss this problem in the context of the novel learning algorithm called *Tacit Learning*[33][34]. Tacit learning is an unsupervised learning scheme based on the above features of biological regulatory systems. One of the most important points of tacit learning is that the learning process and the behavior control process progress in parallel. Roughly defined motions are modified into sophisticated behaviors adapted to the environment through body/environment interactions. All computations for behavior modifications by tacit learning are governed by the autonomous activities of the artificial computational media.

Another feature of tacit learning is the definition of the behavior target. In tacit learning, the behavior target is defined as the constraint conditions that should be achieved independent of the environmental situations. *Survival* is the ultimate constraint condition of tacit learning because all living organisms must achieve this condition independent of the environmental situation. In the case of

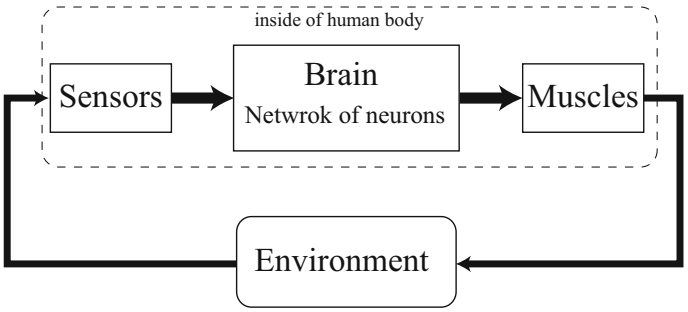

Fig. 1. Conceptual image of biological control loop

the learning of bipedal walking by tacit learning [34], the motions of the swing leg are used as the constraint conditions that could correspond to walking reflexes observed as the instinctive behaviors of newborn babies[35]. The target behavior in tacit learning is therefore strongly related to our instinctive behaviors. The way of specifying the instinctive behaviors for tacit learning is discussed in the following sections.

This chapter is organized as follows: Section 2 explains the features of tacit learning and the way of specifying target behaviors. Section 3 provides typical examples of tacit learning with experimental results obtained using a 2DOF manipulator and discusses the process of acquiring the environmental information based on the precise models of the manipulator and the networks. Section 4 presents the adaptability of tacit learning in terms of the energy consumption, walking rhythm tuning and robustness through the bipedal walking experiments using a 36DOF humanoid robot. Section 5 concludes this chapter.

2 Tacit Learning

The basic idea of tacit learning is the learning of robot behaviors adapted to the environment through body-environment interactions. The robot initially has roughly defined motions consisting of reflexive actions and instinctive behaviors, and tacit learning modifies these motions into sophisticated behaviors adapted to the environment.

2.1 Features of Tacit Learning

The fundamental idea of tacit learning is derived from the features of biological regulatory systems in which all regulations result from the spatial and temporal integration of simple and homogeneous computational media. Based on this characteristic of biological systems, tacit learning is characterized by four features. First, the controller for tacit learning is a network of homogeneous computational media. Learning progresses through accumulating the individual

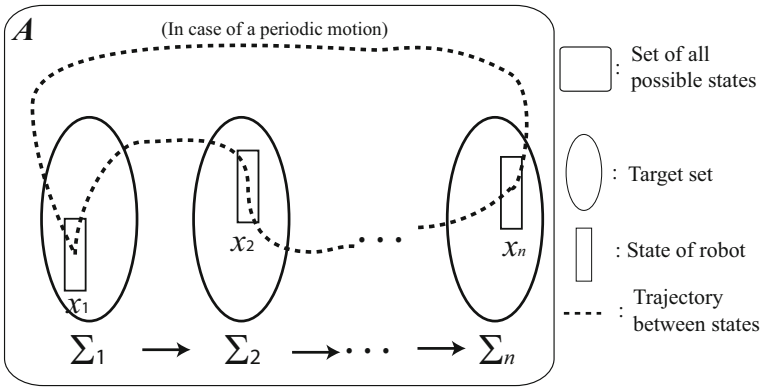

Fig. 2. Conceptual image of a task and its execution

activities of computational media that operate according to innate rules. Second, the sensor-motor connections in the network are organized such that sensor inputs are reduced by motor actions. These innate connections create reflexive actions through body-environment interactions. The combination of these reflexive actions generates primitive motions taking the environmental information into the network. Third, the target behavior is embedded in the network as the constraint conditions that should be achieved independent of the environment. Fourth, no supervising signal is used in the learning process. The learning progresses by accumulating the environmental information inside the network by the activities of the computational media. This learning strategy makes it possible for the behavior control and the behavior adaptation to progress in parallel.

2.2 Definition of Target Behavior for Tacit Learning

Tacit learning should be discussed with specified target behaviors because tacit learning is the way of tuning roughly defined behaviors to sophisticated ones. In the learning of bipedal walking discussed in [36], the specified target behavior was the part of walking locomotion that swung the legs forward alternately. Balance and walking rhythm emerged through tacit learning depending on the condition of the walking surface and the weight of the robot.

We can rigorously define the way of specifying target behaviors as *task* for tacit learning. Let $\boldsymbol{x}$ denote the vector representing the state of a plant to be controlled such that the behavior is described by a transition of $\boldsymbol{x}$ from an initial state to a specified state. For example, the motions of an n-DOF arm are described by the transition of the state of the joint space, $\boldsymbol{x} = [\theta_1\ \theta_2\ \cdots \theta_n]^T$, where, θ_i denotes the angle of Joint i. The target behavior is transition of $\boldsymbol{x}$ from an initial state to a specified state in the state space. These specified states are called *target states*.

State space is usually very large especially in the learning of devices with many degree of freedoms, while the target states frequently involve only a handful of degree of freedoms. Let us take as an example the learning of the motion of picking up an object by using a redundant arm. One of the target states in the motion should be the posture in which the end-effector of the arm reaches the object, which is not unique to redundant arms. In such cases, the target state can be expressed as the set in which every states $\boldsymbol{x}$ can reach to the object. The posture can be chosen from the set depending on the environment such as the position of obstacles. We call the set of the target states a *target set*.

The task for tacit learning is defined as a series of target sets. Figure 2 is a conceptual image illustrated the definition of the task. $\boldsymbol{A}$ and $\boldsymbol{\Sigma_i}$ denote a set of all possible states of $\boldsymbol{x}$ and the target sets, respectively. In the case of bipedal walking, the target sets are defined by the motions of the swing leg. The motions of the supporting leg and other joints are chosen from the target sets. The behaviors of the plant to be controlled are created by choosing trajectories that connect the target sets, and the behaviors may be ones adapted to the environment when appropriate trajectories are chosen. As mentioned in the previous section, we use tacit learning to choose the trajectories for creating adapted behaviors by acquiring environmental information through the reflexive actions.

3 Behavior Learning of 2DOF Manipulator by Tacit Learning

3.1 Network Structures for Tacit Learning

Tacit learning is executed through accumulating the activities of computational media. We proposed artificial computational media [36] whose activities were fundamentally governed by the classical McCulloch-Pitts neuron model[37], which has two states, firing denoted by 1 and rest denoted by 0. The difference between our model and the original McCulloch-Pitts model is the threshold modification to maintain the firing frequency in an appropriate range. That is, our neurons are variable-threshold neurons (VTNs), and each neuron's threshold is increased when the neuron fires and decreased when it rests. The neuron model is described as follows:

$$X(t) = \mathbb{1}(s(t) - \theta(t)) \tag{1}$$

$$\mathbb{1}(u) = \begin{cases} 1 & u \geq 0 \\ 0 & u < 0 \end{cases} \tag{2}$$

$$\theta(t+1) = \theta(t) + \overline{\Delta\theta} X(t) + \underline{\Delta\theta}(X(t) - 1) \tag{3}$$

where $X(t)$, $s(t)$, $\theta(t)$, $\overline{\Delta\theta}$, and $\underline{\Delta\theta}$ denote, respectively, the neuron state at time t, the input to the neuron, the threshold of the neuron, and the values for threshold tuning after firing and rest.

Another activity rule of VTNs is a way of changing the connection weight between them. We extend the Hebbian rule[38] to increase the stability of the

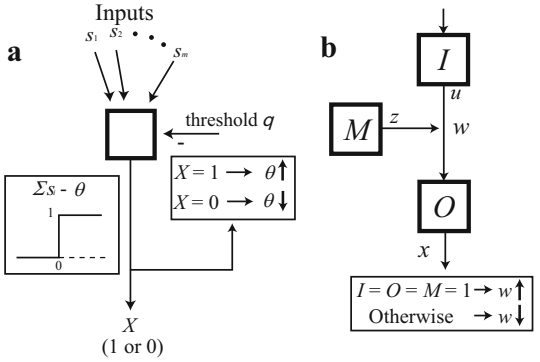

Fig. 3. Activity rules of variable-threshold neuron

network connections. According to the Hebbian rule, the connection weight between two VTNs, I and O, connected as illustrated in Fig. 3 **b** would increase if they fired simultaneously and decrease if they fired separately. We extend this rule by introducing another VTN called a *mediator* (M) so that the connection weight increases if I, O and M are simultaneously fired, otherwise decreases. The extended hebbian rule is described as follows:

$$w(t+1) - w(t) = \overline{\Delta w} x(t)u(t)z(t) + \underline{\Delta w}(x(t)u(t)z(t) - 1) \tag{4}$$

where $x(t)$, $u(t)$, and $z(t)$ correspond, respectively, to the states of O, I and M at t. If both $\overline{\Delta w}$ and $\underline{\Delta w}$ are positive, Eq. (4) is called *potentiation*, and if they are negative, it is called *inhibition*. We take the convention of using a bar to indicate a mediator action of inhibition. Note that both $\overline{\Delta\theta}$ and $\underline{\Delta\theta}$ are positive, while $\overline{\Delta w}$ and $\underline{\Delta w}$ can be negative.

A network of VTNs has strong computational power such as four arithmetic operations, conditioned reflexes, and input accumulation by selecting appropriate threshold patterns[36]. We have developed two networks based on these computational powers. One is an output regulation network, that can find the appropriate threshold pattern of VTNs to constrain the state of an unknown plant to a specified reference, and the other is a self-reference generation network, that can find the threshold pattern that can reduce the quantity of input to the plant by creating a control reference through body-environment interactions.

Block diagrams of the overall network configurations are shown in Fig. 4, where controller C is a network of VTNs that is composed of a serise of VTNs and is called a *cluster* (Fig. 5 **a**). The output from the cluster is the number of firing VTNs. The initial values of all thresholds in the cluster are set such that they are equally distributed in a single band of width $\alpha + \beta$. Here, α and β respectively denote the values for threshold incremental step $\overline{\Delta\theta}$ and decremental step $\underline{\Delta\theta}$ in Eq. (3). Under this assumption on the thresholds, the output from the cluster becomes 0 when the input is smaller than any of the thresholds, and all the VTNs fire when the input is larger than all the thresholds. Actually, the

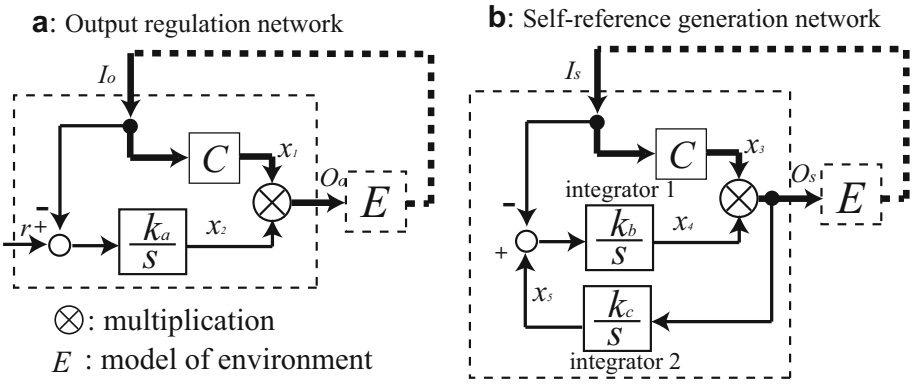

Fig. 4. Block diagrams of **a** output regulation network and **b** self-reference generation network

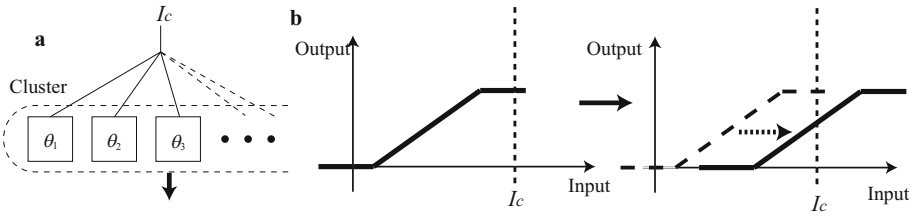

Fig. 5. Features of cluster: **a** Cluster is a group of VTNs that have same inputs. Output from a cluster is the number of firing VTNs. **b** Output reaches a saturation value. Tuning VTN thresholds moves the non-saturated area toward the input.

output from the cluster becomes the saturation system described in Fig. 5 **b**. An interesting feature of the cluster activity is that the non-saturated area moves depending on the output from the cluster to bring the non-saturated area closer to the input value. Tuning the thresholds of VTNs accroding to the rules Eqs. (1)-(3) results in the non-saturated area converging to the value appropriate to the environment. The mathematical expression for the activity of the cluster is given in the Appendix.

As described in Fig. 4 **a**, the output regulation network's output O_o is the value obtained by multiplying the cluster output x_1 by the integrator output x_2, which is the integral value of the difference between the reference r and the input I_o. At the equilibrium state of the output regulation network, the input I_o converges to the reference r and the cluster finds the appropriate threshold patterns in the environment. The control loop shown by the thick lines acts as a reflexive response. The role of this loop is discussed, along with the experimental results, in the next subsection.

In the self-reference generation network, the reference signal in the output regulation network is replaced by the result of integrating output O_s. The O_s should be 0 at the equilibrium state, otherwise the output x_5 from the integrator continues changing. Thus the self-reference generation network can find the appropriate threshold pattern that tunes the output from the network to 0 if the environment allows this. The mathematical expressions of these network activities are given in the Appendix.

3.2 Posture Control of 2DOF Manipulator by Tacit Learning

To experimentally demonstrate the process of creating behaviors by tacit learning, we gave the 2DOF manipulator shown in Fig. 6 **a** the task of making a specified angle in the upper joint, called Joint 2, from the vertically standing posture as shown in Fig. 6 **b** and **c**. In this task, the angle of the lower joint (Joint 1) was not specified but was chosen from the target set by tacit learning. The target set $\boldsymbol{\Sigma}$ of this task is described as follows:

$$\boldsymbol{\Sigma} = \{(\theta_1,\ \theta_2) \mid \forall\theta_1,\ \theta_2 = \theta_d\}, \tag{5}$$

which can be expressed as a line in the Cartesian space of θ_1 and θ_2 illustrated in Fig. 7. Here, θ_1 and θ_2 denote the angles of Joints 1 and 2, and θ_d denotes the desired angle of Joint 2.

As described in Fig. 8 we used the self-reference generation network to control Joint 1 and the output regulation network to control Joint 2. The plant in this configuration can be described as follows:

$$\boldsymbol{M}\begin{bmatrix}\ddot{\theta}_1\\ \ddot{\theta}_2\end{bmatrix} + \boldsymbol{B} = \boldsymbol{U} \tag{6}$$

$$\boldsymbol{Y} = \boldsymbol{C}\boldsymbol{\theta} \tag{7}$$

$$\boldsymbol{M} = \begin{bmatrix} I_1 + m_1a_1^2 + l_1^2m_2 + \zeta + 2\xi\cos\theta_2 & \zeta + \xi\cos\theta_2 \\ \zeta + \xi\cos\theta_2 & \zeta \end{bmatrix} \tag{8}$$

$$\boldsymbol{B} = \begin{bmatrix} -\xi(2\dot{\theta}_1 + \dot{\theta}_2)\dot{\theta}_2\sin\theta_2 + k_1\cos\theta_1 + k_2\cos(\theta_1 + \theta_2) \\ \xi\dot{\theta}_1^2\sin\theta_2 + k_2\cos(\theta_1 + \theta_2) \end{bmatrix} \tag{9}$$

$$\boldsymbol{U} = \begin{bmatrix} u_1\ u_2 \end{bmatrix}^T \tag{10}$$

$$\boldsymbol{Y} = \begin{bmatrix} y_1\ y_2 \end{bmatrix}^T \tag{11}$$

$$\boldsymbol{C} = \begin{bmatrix} 1.0\ 0.1\ 0.0\ 0.0 \\ 0.0\ 0.0\ 1.0\ 0.1 \end{bmatrix} \tag{12}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1\ \dot{\theta}_1\ \theta_2\ \dot{\theta}_2 \end{bmatrix}^T \tag{13}$$

$$\zeta = I_2 + m_2a_2^2 \tag{14}$$

$$\xi = l_1m_2a_2 \tag{15}$$

$$k_1 = (m_1a_1 + m_2l_1)g \tag{16}$$

$$k_2 = m_2a_2g \tag{17}$$

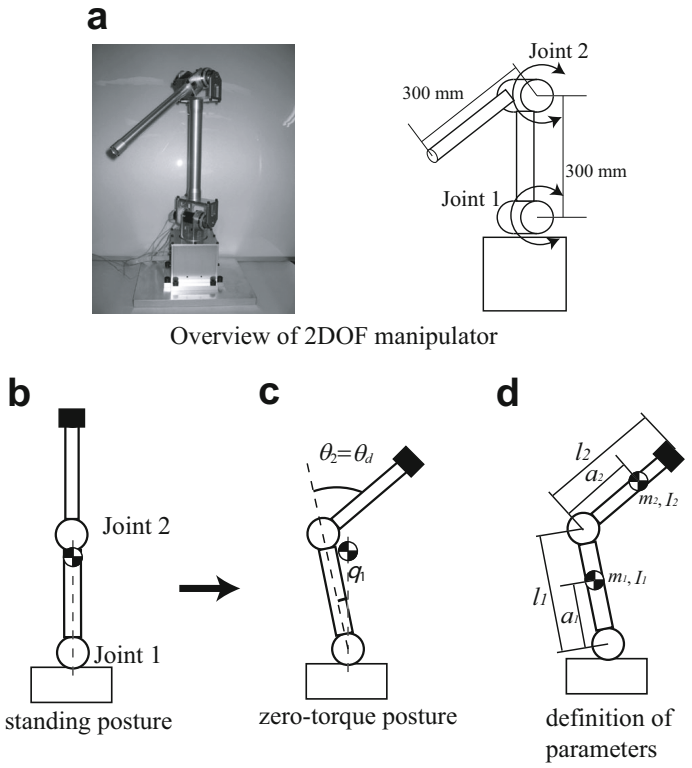

Fig. 6. 2DOF manipulator:**a** Overview of 2DOF manipulator **b** Posture of manipulator standing vertically **c** When the angle of the upper joint is specified and lower joint remains free to be adapted to the environment, the posture in which the center of mass of the manipulator is on a vertical line passing through the attachment point of the lower joint is called the zero-torque posture. **d** Definition of parameters

I_i : Inertia moment of each arm (See Fig.6)
l_i : Length of arm
a_i : Length from joint to center of gravity
m_i : Mass of arm (m_2 includes mass of payload.)

The most interesting feature of this experiment is how the self-reference generation network for Joint 1 finds its values during learning. Figure 9 illustrates the experimental results. A payload with amass of 360g was attached to the top of the manipulator in this experiment. The angle of Joint 2 smoothly converged to the pre-defined reference, which was $\pi/4$ rad in this experiment. Joint 1 was first rotated in the positive direction because the balance of the manipulator was disrupted by the motion of Joint 2. The reflexive action that was mainly

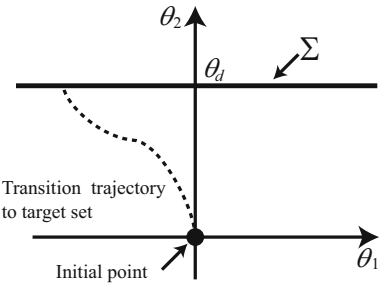

Fig. 7. Representation of target set $\boldsymbol{\Sigma}$ in Cartesian space. The state in $\boldsymbol{\Sigma}$ and the trajectory to the state were chosen through body-environment interaction.

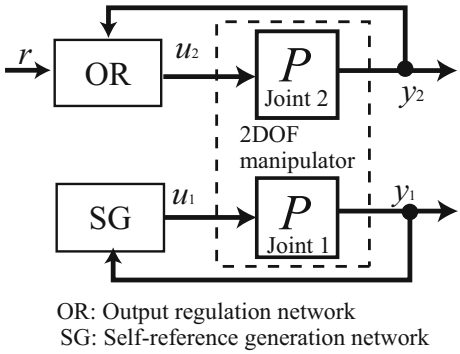

Fig. 8. Controller configuration for 2DOF manipulator control: The output regulation network was used to control the upper joint (Joint 2) and the self-reference generation network was used to control the lower joint (Joint 1).

controlled by the loop shown by the thick lines in Fig. 4 **b** appeared immediately after this rotation of Joint 1. The reflexive action of Joint 1 stimulated other network loops in Fig. 4 **b**. As shown in Fig. 9 **c**, the angle of Joint 1 finally converged to -0.26 rad, where the manipulator could be kept balanced without applying torque to Joint 1 . Through tacit learning, the self-reference generation network worked to find a *zero-torque posture* in which no torque was applied to Joint 1.

The equilibrium point of the self-reference generation network is described as follows:

$$y_1(t) = x_5(t),\ x_4(t) = 0.0,\ u_1(t) = 0.0,\ x_3(t) = \frac{N\beta}{\alpha + \beta}, \tag{18}$$

where N denotes the number of VTNs in the cluster. The details on the derivation of the equilibrium point can be found in the Appendix. Under the mechanical limitation of Joint 1 which was $-\pi < \theta_1 < \pi$, the angle of Joint 1 at the

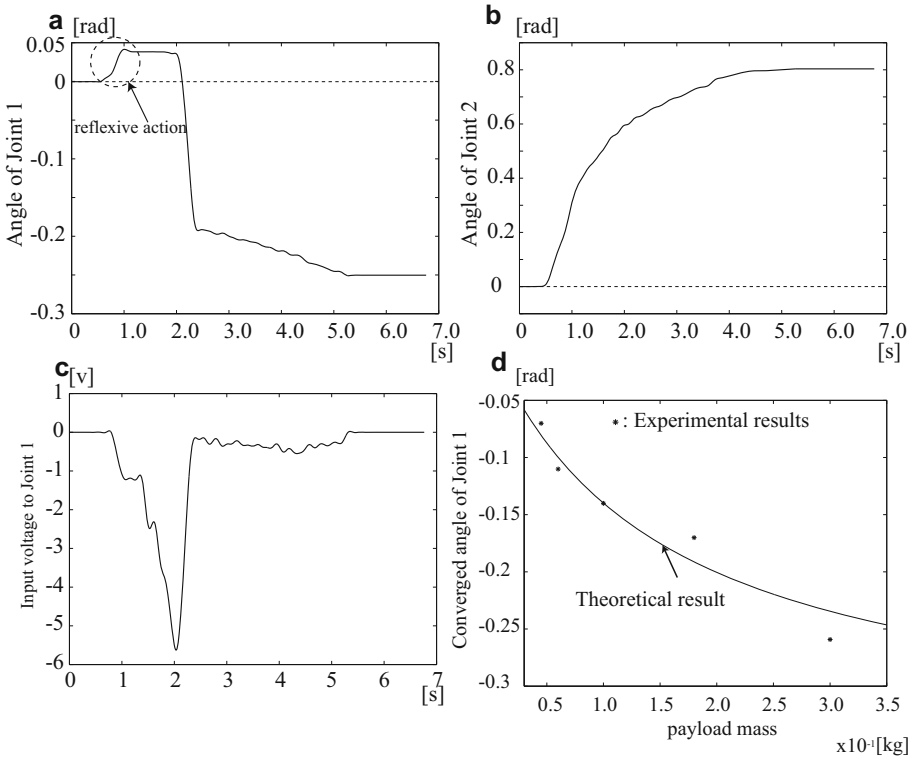

Fig. 9. Experimental results:**a** time history of angle of Joint 1 **b** time histories of angle of Joints 2 **c** time history of voltage working on Joint 1 **d** converged angle of Joint 1 as a function of payload mass

equilibrium point becomes

$$\theta_1 = \tan^{-1}\left(\frac{k_2 \cos\theta_d + k_1}{k_2 \sin\theta_d}\right), \tag{19}$$

which corresponds to a zero-torque posture. The reflexive action led the angle of Joint 1 to the above equilibrium angle, which was determined by the environment, the body parameters of the manipulator, and the specified angle of Joint 2.

The reflexive action of Joint 1 was created by body-environment interaction, which in this case was the loss of balance caused by gravitational force acting on the manipulator body. Thus when the body parameters were changed, the motion of Joint 1 automatically changed. $*$ in Fig. 9 **d** shows the converged angles of Joint 1 obtained with various payload masses. Without information to the network about the changes, the angle of Joint 1 converged to the neighborhood of the equilibrium point calculated using Eq. (18) (solid line in Fig. 9 **d**).

3.3 Emergence of Reciprocating Motion

Next, we gave alternative θ_d to Joint 1, $\theta_1 = 0$ and $\theta_d = \pi/4$ as described in Fig. 6 **b** and **c**. The target sets of these two postures are defined as follows:

$$\boldsymbol{\Sigma}_1 = \{(\theta_1,\ \theta_2) \mid \theta_1 = 0.0,\ \theta_2 = 0.0\} \tag{20}$$

$$\boldsymbol{\Sigma}_2 = \{(\theta_1,\ \theta_2) \mid \forall\theta_1,\ \theta_2 = \pi/4\} \tag{21}$$

Here, the target set Σ_1 is reduced to a single point representing the standing posture without applying torques to both joints. When the manipulator was moved from Σ_1 to Σ_2, the controllers described in Fig. 4 were used. When it moved in reverse from Σ_2 to Σ_1, both joints were controlled by the output regulation networks. The control direction was switched when the states of the manipulator were sufficiently close to Σ_1 and Σ_2.

Experimental results obtained with a 306 g payload are shown in Fig. 10. The periodic motion appeared after several reciprocating motions, in which zero-torque postures in Σ_2 were chosen. This result implies that the periodic motion was created between the equilibrium points in Σ_1 and Σ_2.

4 Learning of Bipedal Walking by Tacit Learning

This section describes experiments in which we applied tacit learning to a much more complex bipedal walking problem and investigated how well the behaviors created by tacit learning were adapted to the environment.

4.1 Definition of Task for Bipedal Walking

We used the 36DOF humanoid robot described in Fig. 11 in the experiments. As discussed in Section 2, the motions of putting the legs forward alternately were used . Actually, we took as the target sets four postures in one step (Table 1). In target set, we specified the desired angles of some joints on the swing leg and didn't specify the motion of the supporting leg like Joint 1 in the 2DOF manipulator experiments. The output regulation networks were used to control the specified joints and the self-reference regulation networks were used for the other joints.

4.2 Experiments on Bipedal Walking

The reference values for the specified joints in the experiments are summarized in Table 1. The target set was switched to the next one when the specified angles converged to the references. To create periodic motion, Σ_8 and Σ_1 were connected.

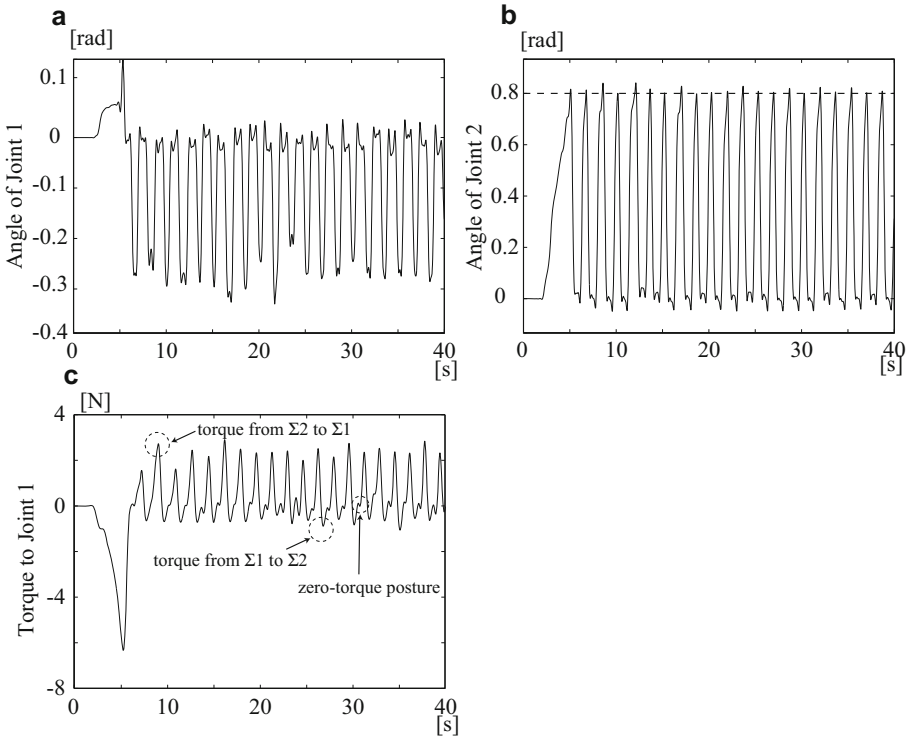

Fig. 10. Experimental results of 2DOF manipulator posture control:**a**, **b** and **c** show time histories of the angles of Joints 1 and 2 and of the torque applied to Joint 1, respectively. Direction of the trajectory was switched when the state converged to the constraint conditions.

Movies of the experiments are on [39]. In the initial state, the robot fell down at the initial state even though its legs moved forward. After about 10 minutes, the motion of the supporting leg was tuned and the robot kept walking. Figure 12 describes the lateral angle of the hip joint (Jr 6 in Fig. 11) before and after walking was learned. The motion of the joint that rotated randomly during the initial couple of minutes gradually became periodic, and eventually periodic motion emerged after 10 minutes.

Figure 13 describes the trajectories of Jr 6 when the state moved from $\boldsymbol{\Sigma_2}$ to $\boldsymbol{\Sigma_3}$. The broken lines represent the trajectories that were used before learning was complete. These lines appeared in the first 3 minutes. The solid lines are the trajectories after the robot became able to walk continuously, which occurred in the final 2 minutes. The trajectory modification from the broken to the solid lines happened during the process of searching for the equilibrium point through the reflexive actions, which was the same process as that for Joint 1 in the 2DOF manipulator experiments discussed in the previous section. We observed similar convergences of trajectories for other unspecified joints.

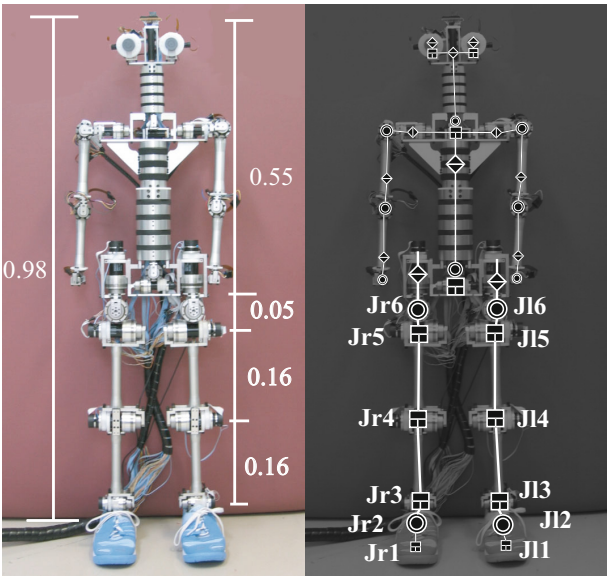

Fig. 11. Overview of 36DOF humanoid robot (distance are specified in meters)

Table 1. Target posture for walking and specified angles

Target set (key frame)	Description	Specified DOF (Value in experiment [rad])
$\sum_1$	Balance on Right Leg	Jr 6 (0.08) Jl6 (-0.08)
$\sum_2$	Left Leg Up	Jl4 (0.4) Jl5 (0.2)
$\sum_3$	Left Leg Down	Jl4 (0.0) Jl5 (0.2)
$\sum_4$	Waiting after Left Leg Step	-
$\sum_5$	Balance on Left Leg	Jr6 (-0.08) Jl6 (0.08)
$\sum_6$	Right Leg Up	Jr4 (-0.4) Jr5 (-0.2)
$\sum_7$	Right Leg Down	Jr4 (0.0) Jr5 (-0.2)
$\sum_8$	Waiting after Right Leg Step	-

4.3 Adaptability of Bipedal Walking to Environment

Our interest is how well the created walking gait was adapted to the environment, and here we discuss the following three aspects of this question.

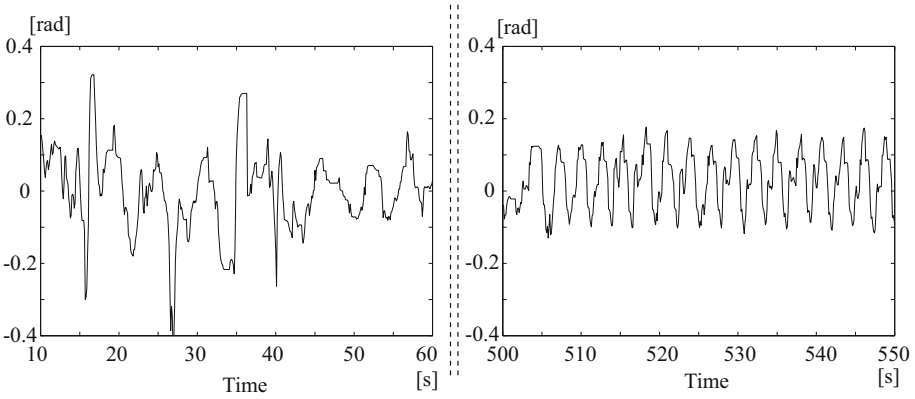

Fig. 12. Time histories of joint Jr6: In the first 60 s the trajectories rotated randomly and caused the robot to fall down. After 500 s from the start of the learning, however, regular rhythm appeared and the robot kept walking.

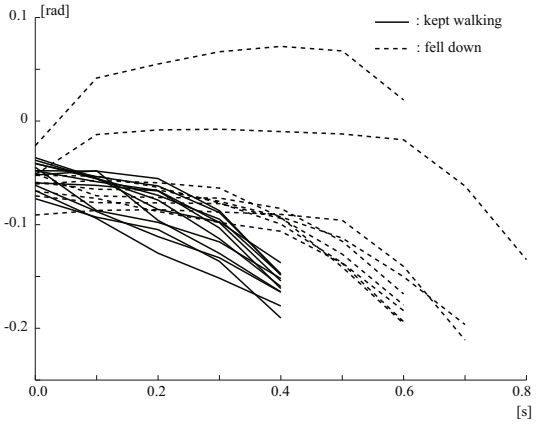

Fig. 13. Trajectories of joint Jr6 when the state moves from $\boldsymbol{\Sigma_2}$ to $\boldsymbol{\Sigma_3}$: The broken lines are the trajectories that caused the robot to fall down after the state moved on the trajectories. The solid lines are the trajectories where the robot kept walking. Because of the data sampling time during the experiments, 0.1 s, all the solid lines are stopped at 0.4 s. This implies that the error in walking pace is within 0.1 s.

The first is the efficiency of the walking gait, which is one of the most important indexes of a walking gait's adaptation to the environment[40]. It is natural to think that better efficiency implies a waking gait more adapted to the environment. We use the following index to evaluate efficiency[41]:

$$walkingsurfacesE = \frac{\text{energy comsumption}}{(\text{mass of the robot}) \times (\text{traveled distance})}. \qquad (22)$$

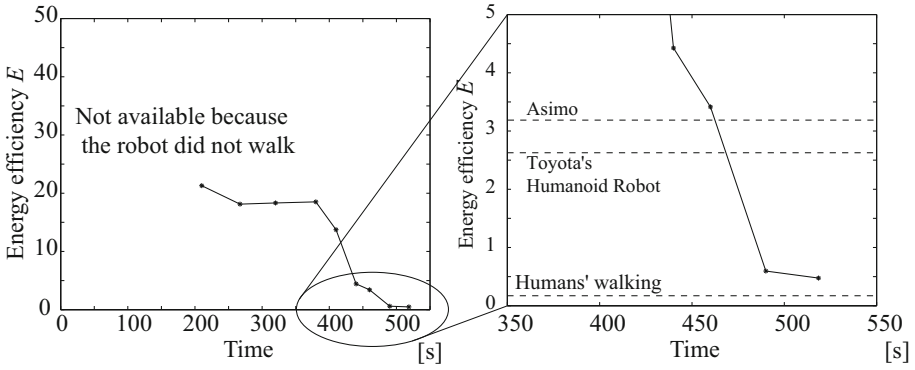

Fig. 14. Time history of energy efficiency E described in Eq. (22) during learning bipedal walking: At around 300s from the start learning, the energy efficiency was much worse than that for the fully controlled humanoid robot. However, around 500s, when walking rhythm appeared, efficiency improved greatly and the robot walked with efficiency near that of humans

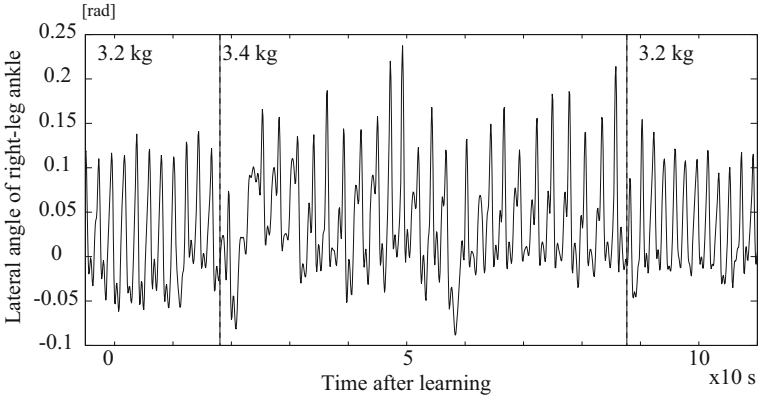

Fig. 15. Changes in walking rhythm depending on weight: The rhythm was tuned slower when the weight increased and *vice versa* without any explicit information about the change in weight.

Figure 14 describes the time history of this index during the learning of walking. The efficiency of a human's walking and of two other fully controlled humanoid robot's walking is also illustrated in Fig. 14[41][42]. The results demonstrate that the efficiency of our robot improved as the learning progressed. At the final stage of learning, it became less than one-fifth of that of other fully controlled humanoid robots and was almost the same value as that of a human walking. This remarkably high level of efficiency was achieved by reducing power consumption by maintaining balance without torque during walking. This walking style is

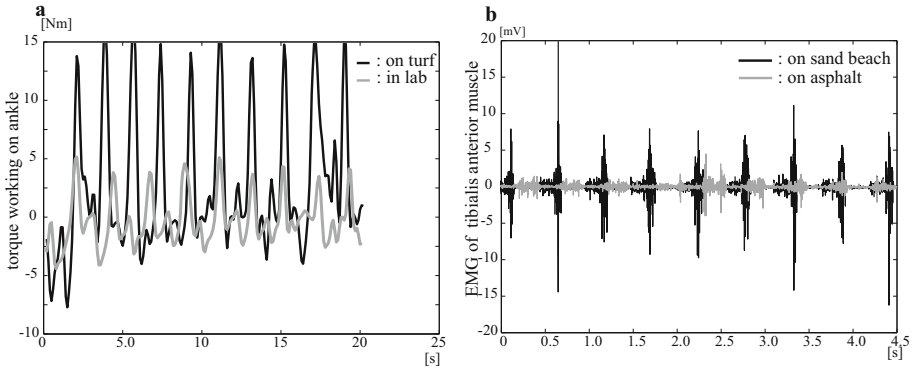

Fig. 16. Walking-surfaces-dependent difference in torque working on the ankle: **a** Much larger torque to the ankle joint was used when the robot walked on turf than on a hard flat surface. **b** Myoelectric potential of the tibialis anterior muscle becomes larger when human walks on a sandy beach that on asphalt.

similar to that of humans and corresponds to the zero-torque posture in the 2DOF manipulator experiments.

The second aspect we discussed here is the autonomous changes in the walking rhythm. We did not set any time-dependent parameters in the experiment: the periodic walking rhythm emerged through body-environment interactions. Thus when the body parameters and/or the environment were changed, the rhythm was automatically changed as described in Fig. 15, which shows what happened when the weight of the robot was changed abruptly after it had learned the walking. Without any explicit information about the change in weight, the rhythm was tuned slower when the weight increased and *vice versa*. These modifications are reasonable adaptation to the weight changes.

The final aspect we consider here is how the torque working on the ankle differs depending on the walking surface. As we can see from the movies on [39], our method succeeded in creating walking on natural turf, not just on the flat and hard surface in the laboratory. Figure 16 **a** describes the time histories of the torque working on the ankle after the learning of the walking. The results indicate that the controller required much larger torque for walking on turf than for walking in the laboratory. We can observe similar changes of torque in our walking. Figure 16 **b** shows electromyogram (EMG) data from the tibialis anterior muscle, which controls the angle of the ankle, obtained when a human walked on asphalt and on a sandy beach. As seen in Fig. 16 **b**, we empirically know that we use much more power to keep balance while walking on a sandy beach than on asphalt. Thus, the emergence of appropriate torque depending on the surface conditions in our experiments implies that body-environment interactions created a walking gait that was adapted to the environment.

The observations on the above three aspects imply that the behaviors created by tacit learning are not just adapted to the environment but also share many features with biological behaviors. We believe that these similarities were derived from the features of the activity rules of VTNs in the network, which changed their thresholds to reduce their outputs when the environmental inputs increased. In tacit learning, the environmental information was analyzed according to these features. The processes that can reduce the outputs carrying out a specified task tend to save energy in the environment. This would also be important for biological systems to increase the chance of survival in a natural environment.

5 Conclusion

The notion of tacit learning has been introduced to develop artificial control systems with remarkable adaptability to unpredictable environmental changes. The fundamental computational algorithm for tacit learning is based on the feature of biological regulatory systems in which all regulations result from the spatial and temporal integration of homogeneous computational media that act subject to innate rules. A network of the homogeneous computational media that connects the sensors and motors in proper ways is of great advantage in orchestrating the flow of heterogeneous environmental information. We developed networks of artificial computational media and used them to control a humanoid robot.

The experimental results demonstrated that the reflexive actions originating from the innate sensor-motor connections in the network changed primitive motions into sophisticated behaviors adapted to the environment. Even small changes in the environment influenced the learning results because the reflexive actions were caused by body-environment interactions. The three observations using 36DOF humanoid robot discussed in Section 4 verified the high adaptability of tacit learning in terms of gait generation, power consumption, and robustness.

The environmental information taken into the network through the reflexive actions played the roles of supervising signals for learning. This learning scheme is strongly associated with the notion of *affordance*[43], which is recognized as the key factor in cognition and intelligence. In our case, the environmental information was mainly used to create the motions of the joints without concrete references and led to the adapted behaviors. The creation of meaningful behaviors from purposeless actions by using the environmental information should be an essential process to establish adaptation and intelligence in man-made machines. This feature of tacit learning could open up innovative fields that can accomplish high levels of adaptability and artificial intelligence, and can provide a better understanding of biological systems.

The analysis of the tacit learning process in this chapter is limited to several behaviors. To the learning of multiple behaviors, we need further theoretical analysis to clarify the role of the computational media, reflexive actions, and the process of searching equilibrium to extend tacit learning . The way of the specification of the target sets, which in this chapter were given by the specified angles

of some joints, also warrants further discussion. To create behaviors that enable robots to act as real partners of human beings, behavioral objectives should be much more abstract like the eventual goal of all living organisms, *survival*. We are now on the way to creating multiple behaviors through interactions with more complicated environments.

Appendix: Mathematical expression of proposed computational media and controllers

The cluster described in Fig. 5 **a** is composed of a series of VTNs with a common input. The output from the cluster is the number of firing VTNs. When the input to the cluster is sufficiently large or small, all VTNs fire or none do. Thus, the cluster becomes the saturation system. Under some assumption on the distribution of the thresholds, we can prove the following input/output relationship of the cluster[33]:

$$O(t) = \mathrm{SAT}_{N,\gamma}(\eta(t)) \tag{23}$$

$$\Theta(t) = \Theta(t-1) + \frac{\gamma}{N} O(t-1) - \beta \tag{24}$$

$$\eta(t) = I(t) - \Theta(t) \tag{25}$$

$$\gamma = \alpha + \beta \tag{26}$$

Here, $O(t)$, N, $\Theta(t)$, $I(t)$, α, and β denote the output from the cluster, the number of VTNs in the cluster, the average of the thresholds of all VTNs, the input to the cluster, the value for the threshold incremental step $\overline{\Delta\theta}$, and that for the decremental step $\underline{\Delta\theta}$ in Eq. (3), respectively. $\mathrm{SAT}_{a,b}(x)$ is the saturation function described as follows:

$$\mathrm{SAT}_{a,b}(x) = \begin{cases} a & x > \frac{b}{2} \\ \left[\frac{a-1}{b}\left(x + \frac{b}{2}\right)\right] - 1 & |x| \le \frac{b}{2} \\ 0 & x < -\frac{b}{2} \end{cases} . \tag{27}$$

Here, $[*]$ denotes the floor function that expresses the next smallest integer of $*$.

When the input is a constant value, the output from the cluster converges to the following value that is independent of the input value:

$$O(t) = \frac{N\beta}{\gamma}. \tag{28}$$

By using these equations, the output regulator and the self-reference generator described in Fig. 4 are described by the following hybrid systems:

Output regulator:

$$O_o(t) = x_1(t)x_2(t) \tag{29}$$

$$x_1(t) = \mathrm{SAT}_{N,\gamma}(\eta_o(t)) \tag{30}$$

$$x_2(t) = -k_a \int_0^t (I_o(\tau) - r)d\tau \tag{31}$$

$$\Theta_o(t) = \Theta_o(t-1) + \frac{\gamma}{N} x_1(t-1) - \beta \tag{32}$$

$$\eta_o(t) = I_o(t) - \Theta_o(t) \tag{33}$$

Self-reference generator:

$$O_s(t) = x_3(t)x_4(t) \tag{34}$$

$$x_3(t) = \mathrm{SAT}_{N,\gamma}(\eta_s(t)) \tag{35}$$

$$x_4(t) = -k_b \int_0^t (I_s(\tau) - x_5(\tau))d\tau \tag{36}$$

$$x_5(t) = k_c \int_0^t O_s(\tau)d\tau \tag{37}$$

$$\Theta_s(t) = \Theta_s(t-1) + \frac{\gamma}{N} x_3(t-1) - \beta \tag{38}$$

$$\eta_s(t) = I_s(t) - \Theta_s(t) \tag{39}$$

x_i : Signals in controller (See Fig.4)
I_o and O_o : Inputs and outputs of plants in Output Regulator
I_s and O_s : Inputs and outputs of plants in Self − reference Generator
Θ_o and Θ_s : Average of thresholds in clusters

From the above equations, the values of parameters at equilibrium states are derived as follows:

$$x_1 = \frac{N\beta}{\gamma},\ I_o = r,\ I_s(t) = x_5(t),\ x_4(t) = 0.0,$$
$$u_1(t) = 0.0,\ x_3(t) = \frac{N\beta}{\alpha + \beta} \tag{40}$$

The values of x_2, O_o, I_s depend on the environment in which these controllers are used.

In the experiments in this chapter, we used the following values:

$$N = 100,\ \alpha = 1.0 \times 10^{-2},\ \beta = 1.0 \times 10^{-3}, \Theta_i(0) = -0.7,$$

$$k_a = 1.0 \times 10^{-2}, k_b = 1.0 \times 10^{-2}, k_c = 1.0 \times 10^{-3}. \tag{41}$$

References

1. Bear, M.F., Connors, B.W., Paradios, M.A.: Neuroscience: Exploring the brain, 3rd edn. Lippincott Williams and Wilkinse (2006)
2. Keener, J.: Mathematical Physiology. Springer (1998)
3. Tanaka, R.J., Kimura, H.: Mathematical classification of regulatory logics for compound environmental changes. Journal of Theoretical Biology 251, 363–379 (2008)
4. Janeway, C.A., Travers, P., Walport, M., Shlomchik, M.J.: Immunobiology: The Immune System in Health and Disease, 5th edn. Garland Science, New York (2001)
5. Shimoda, S., Yoshihara, Y., Fujimoto, K., Yamamoto, T., Maeda, I., Kimura, H.: Stability analysis of tacit learning based on environmental signal accumulation. In: Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (2012)
6. Hodgkin, A.L., Huxley, A.F.: Currents carried by sodium and potassium ions through the membrane of the giant axon of *loligo*. Journal of Physiology 116, 449–472 (1952)
7. Markram, H., Tsodyks, M.: Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature 382, 807–810 (1996)
8. Phares, G.A., Antzoulatos, E.G., Baxter, D.A., Byrne, J.H.: Burst-Induced Synaptic Depression and Its Modulation Controbute to Information Transfer at Aplysia Sensorimotor Synapses: Empirical and Computational Analyses. The Journal of Neuroscience (23), 8392–8401 (2003)
9. Shepherd, G.M.: The synaptic organization of the brain. Oxford University Press (2003)
10. Jacob, F., Monod, J.: Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology 3, 318–356 (1961)
11. Abbott, L.F., Varela, J.A., Sen, K., Nelson, S.B.: Synaptic Depression and Control Gain Control. Science (275), 220–224 (1997)
12. Cook, D.L., Schwindt, P.C., Grande, L.A., Spain, W.L.: Synaptic depression in the localization of sound. Nature 421, 66–70 (2003)
13. Fortune, E.S., Rose, G.J.: Short-term synaptic plasticity as a temporal filter. Trends in Neurosciences 24(7), 381–385 (2001)
14. Castro-Alamancos, M.A.: Different temporal processing of sensory inputs in the rat thalamus during quiescent and information processing states in vivo. Journal of Physiology (539), 567–578 (2002)
15. Eytan, D., Bernner, N., Marom, S.: Selective adaptation in networks of cortical neurons. Journal of Neuroscience (23), 9349–9356 (2003)
16. Castellucci, V.F., Pinsker, H., Kupfermann, I., Kandel, E.R.: Neuronal mechanisms of habituation and dishabituation of the gill-withdrawal reflex in Aplysia. Science 167, 1745–1748 (1970)
17. Grade, L.A., Spain, W.J.: Synaptic Depression as a Timing Device. Physiology (20), 201–210 (2005)
18. Kawato, M., Furukawa, K., Suzuki, R.: A hierarchical network model for motor control and learning of voluntary movement. Biological Cybernetics 57, 169–185 (1987)
19. Minsky, M.L., Papert, S.A.: Perceptron. MIT Press (1969)
20. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature 323(9) (1986)
21. Kuniyoshi, Y., Yorozu, Y., Suzuki, S., Sangawa, S., Ohmura, Y., Terada, K., Nagakubo, A.: Emergence and development of embodied cognition: A constructivist approach using robots. Progress in Brain Research 164, 425–445 (2007)

22. Barto, A.G., Sutton, R.S., Anderson, C.W.: Neuronlike adaptive elements that can solve difficult learning control problems. IEEE Transactions on Systems, Man, and Cybernetics 13, 834–846 (1983)
23. Doya, K.: Reinforcement learning in continuous time and space. Neural Computation 12, 219–245 (2000)
24. Tedrake, R., Zhang, T.W., Seung, H.S.: Stochastic policy gradient reinforcement learning on a simple 3d biped. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (2004)
25. Holland, J.H.: Adaptation in natural and artificial systems. MIT Press (1992)
26. Wang, H., Yang, S., Ip, W.H., Wang, D.: Adaptive primal-dual genetic algorithms in dynamic environments. IEEE Transactions on Systems, Man, and Cybernetics Part B 39, 1348–1361 (2009)
27. Astrom, K.J., Wittenmark, B.: Adaptive control. Addison Wesley (1989)
28. Slotine, J.E., Li, W.: Applied Nonlinear Control. Prentice Hall (1991)
29. Brooks, R.A.: A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation 2, 12–23 (1986)
30. Brooks, R.A.: New approaches to robotics. Science 253, 1227–1232 (1991)
31. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man, and Cybernetics SMC-3, 28–44 (1973)
32. Juang, J.G.: Fuzzy neural network control CMAC of a biped walking robot. IEEE Transactions on Systems, Man, and Cybernetics Part B 30(4), 594–601 (2000)
33. Shimoda, S., Kimura, H.: Bio-mimetic Approach to Tacit Learning based on Compound Control. IEEE Transactions on Systems, Man, and Cybernetics-Part B 40(1), 77–90 (2010)
34. Shimoda, S., Yoshihara, Y., Kimura, H.: Adaptability of tacit learning in bipedal locomotion. IEEE Transactions on Autonomous Mental Development 5(2), 152–161 (2013)
35. Forssberg, H.: Ontogeny of human locomotor control. I. Infant stepping, supported locomotion and transition to independent locomotion. Exp. Brain Res. 57(3), 480
36. Shimoda, S., Kimura, H.: Neural Computation Scheme of Compound Control: Tacit Learning for Bipedal Locomotion. SICE Journal of Control, Measurement, and System Integration 1(4), 275–283 (2008)
37. McCulloch, W.S., Pitts, W.: A logical calculus of the idea immanent in nervous activity. Bull. Math. Biophys. 5, 115–133 (1943)
38. Hebb, D.O.: The organization of behavior. Wiley, New York (1949)
39. `http://btcc.nagoya.riken.jp/biologic/movies.html`
40. Sockol, M.D., Raichlen, D.A., Pontze, H.: Chimpanzee locomotor energetics and the origin of human bipedalism. PNAS 104(30), 12265–12269 (2007)
41. Collins, S., Ruina, A., Tedrake, R., Wisse, M.: Efficient bipedal robots based on passive-dynamic walkers. Science 307, 1082–1085 (2005)
42. `http://workd.honda.com/asimo/`
43. Gibson, J.J.: The ecological approach to visual perception (new edition). Psychology Press (1986)