10

Developmental Learning with Behavioral-Mode Tuning by Carrier-Frequency Modulation

We realize a so-called developmental learning with which a motion-control system learns multiple tasks similar to each other, or advanced ones, incrementally and efficiently by tuning its behavioral mode. The system is based on a coherent neural network whose carrier frequency works as a mode-tuning parameter. The coherent neural network is a class of the complex-valued neural networks. As presented in the previous chapters, we can modulate the behavior of the coherent neural network, such as learning and processing, by changing the carrier frequency. We make the carrier frequency represent the internal mode of the system, and utilize the carrier frequency as the key to realize the developmental learning. In this chapter, we consider two tasks related to bicycle riding. The first is to ride as temporally long as the system can before it falls down (Task 1). The second is an advanced one, i.e., to ride as far as possible in a certain direction (Task 2). We compare developmental learning to learn Task 2 after Task 1 with the direct learning of Task 2. Experiments demonstrate that the developmental learning enhances the efficiency in learning in total. We confirm the effectiveness of the developmental learning utilizing the carrier frequency as the mode-tuning key in the coherent neural network.

10.1 Development, Context Dependence, Volition, and Developmental Learning

Development is an important concept in the science to understand human beings. It has been widely studied in various fields such as cognitive science, psychology and neuroscience. In robotics, development is also expected to play an important role in various applications. For example, Asada et al. [211] proposed cognitive developmental robotics. If we regard a learning process as a search for an appropriate system state, we can interpret the developmental learning as follows. First we begin with a simple and dimensionally small

A. Hirose: Complex-Valued Neural Networks, SCI 400, pp. 151–162.
springerlink.com

state, and then we increase the search dimension, if needed, to realize efficient learning.

The search dimension is often so large in solving real-world problems that developmental learning will be crucial to successful learning or self-organization workable in a realistically short time. The dimension-increasing procedure is also compared to a situation that, for example, a mother will first present a most basic goal to a child and, afterward, give him or her gradually advanced tasks one after another.

What condition is required for the system to realize such developmental learning? Among others, the most important is an appropriate structure that increases the effective number of internal states[1] or the dimension of them efficiently. To be simple, the system should generally prepare a small set of sufficient and effective parameters or variables incrementally. Thereby, in the parameter enlargement, it is significantly important that the parameter increment leads firmly to an expansion of behavior. At this point, the use of a coherent neural network (CNN) is promising, which we describe later.

Regarding the preparation and control of internal states in artificial neural networks, we have several proposals to utilize internal-mode modulations to extend learning and self-organization for the emergence of context-dependent behavior [212]. The context dependence is a behavioral feature explained as follows. For example, assume that people ask you, "What do you like?" If they were talking about sports, then you may answer, "I like tennis." However, if they were discussing swimming, you may answer, "I like backstroke." In such a way, we catch the course of the talk, and respond accordingly. This is one of the context-dependent behavior. Such behavior is regarded as emergence of volition. In other words, we possess a situation-dependent direction or intention, i.e., internal state, inside us, and decide what to do based on it.

In PATON proposed by Omori et al. [213],[214], the behavior of recognition and association in an associative memory system is controlled by a context-dependent switch. In motion control, Wolpert and Kawato [215],[216] prepared multiple neural-network modules. In their system, each output is weighted by a "responsibility coefficient" determined by the closeness of the tentative output value of each module to a desirable one. The coefficient is also used effectively as a weight in the learning process. Then the outputs are weighted and summed to yield a total output signal of the neural system. In such a manner, the system consistently learns an appropriate motion control, and then processes input sensory signals properly. Hartono and Hashimoto [217] also reported the successful introduction of annealing in the module-output integration.

Such switching and weighting-and-integrating methods increase the variety of neural states in learning and processing. In the extension, a crucial

[1] The internal state mentioned here is NOT the neuron's internal state described in Section 2.3 or 4.1.1, but a parameter existing in the network and determining the neural-network's behavior. By changing it, we realize the modulation of mood or intent of the network.

characteristic is again how flexible and effective the system can change its behavior. Simultaneously, a smooth behavioral variation, i.e., the generalization, is also an indispensable characteristic for natural neural processing.

We can expect that developmental learning will also be realized on the basis of CNNs using carrier frequency modulation for the behavioral mode tuning. In general, the CNN has a large freedom of tuning and a flexible generalization characteristic in its behavior by utilizing the complete-orthogonal property of the trigonometric basis functions used ($\cos\theta$ and $\sin\theta$, or $\exp[i\theta]$), for example, $e^{i2\pi f\tau_{ji}} = \cos 2\pi f\tau_{ji} + i\sin 2\pi f\tau_{ji}$ in the complex-valued Hebbian rule in (4.47) and (4.48) [208], [218]. In other words, the summation of a set of weighted sinusoidal curves is potentially capable of yielding a large variety of functions [204],[56].

In this chapter, we present a developmental learning architecture based on the CNN with carrier-frequency modulation for behavioral mode tuning [219]. The developmental learning is also regarded as a short-time growth. First, the network learns a certain task. Then, it learns a similar or advanced task quickly by utilizing the skill obtained previously. We consider bicycle riding. The first task is to ride a bicycle as temporally long as possible (Task 1), while the second and advanced one is to ride as far as possible in a certain direction (Task 2). This procedure is a class of developmental learning, though the situation is very simple. We compare the performance of developmental learning with that of the direct learning of Task two.

Note that the basic idea is already presented in Chapters 8 and 9 where we describe lightwave neural networks whose behavior is dependent on the optical carrier frequency. The same framework realizes a developmental learning. Additionally, in the process, the system finds the best frequency by itself. In this sense, the developmental learning is realized by self-organization.

10.2 Neural Construction and Human-Bicycle Model

Figure 10.1 shows a neuron in the coherent neural network. The input signal x_m, output signal y_n and weight w_{nm} are all complex numbers and composed of amplitude and phase. We adopt an amplitude-phase-type neuron activation function, which is introduced in Section 3.3.5, expressed in terms of the complex-valued input summation $s_n \exp[i\beta_n]$ with amplitude s_n, phase β_n and $i \equiv \sqrt{-1}$ as

$$s_n \exp[i\beta_n] \equiv \sum_m w_{nm} x_m \tag{10.1}$$

$$y_n = A \tanh(g s_n) \exp[i\beta_n] \tag{10.2}$$

where A and g (real numbers) denote saturation amplitude and small-signal gain that determines unsaturated gain, respectively. The function transforms

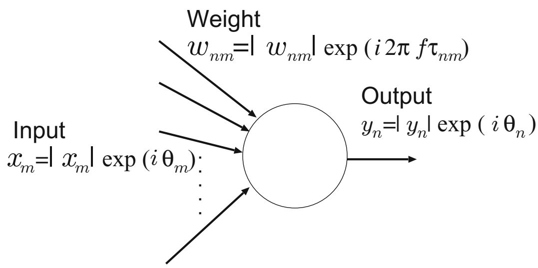


Fig. 10.1 Amplitude-phase-type complex-valued neuron in coherent network.

the signal amplitude in a saturation manner, just like the real-valued sigmoid function does, while it leaves the signal phase unchanged. The operation is a natural extension of the real-valued sigmoidal activation function.

A set of the neurons form a CNN with a carrier frequency, f, that works as the mode parameter. The neural connection weight w_{nm} is expressed by the connection amplitude (transparency) $|w_{nm}|$, delay time τ_{nm} and the carrier frequency f common to all the weights as

$$w_{nm}(f) = |w_{nm}| \exp[i2\pi f \tau_{nm}] \tag{10.3}$$

Therefore, the behavior of the coherent neural network depends on f according to (10.1)-(10.3). As mentioned above, we use this carrier frequency as the modulation parameter of the behavioral mode. If we fix the parameter value f, the behavioral mode is also fixed, whereas if we release it free to move to an optimal point self-organizingly, then the network learns and processes properly with the optimal parameter different from the previous one. A context-dependent behavior is also expected to emerge with this dynamics.

Figure 10.2 shows the construction of the coherent neural network interacting with a bicycle. It is a single-layered feedforward network. Variables are explained below in relation to human-bicycle model. The human-bicycle physical model is shown in Fig.10.3. We have variables such as handlebar azimuth ϕ, bicycle velocity v, wheel torque T, human rolling angle relative to bicycle σ, rolling angle of the total center of gravity of human and bicycle α. We have developed a mechanics simulator which is similar to that used in the study of walking. Figure 10.4 presents a window capture. The $x-y$ section shows a bicycle (larger box) and a human (smaller one) projected on the ground, while the $y-z$ and $x-z$ sections present their elevations. The angle-of-roll section illustrates their rear view of the rolling angle. The curves on the right-hand side show time evolutions of the rolling angle of the total human-bicycle gravity center α, handlebar azimuth ϕ, bicycle velocity v and wheel torque T.

The above variables are shown also in Fig.10.2, together with another variable γ which stands for the azimuth of the bicycle running direction. The

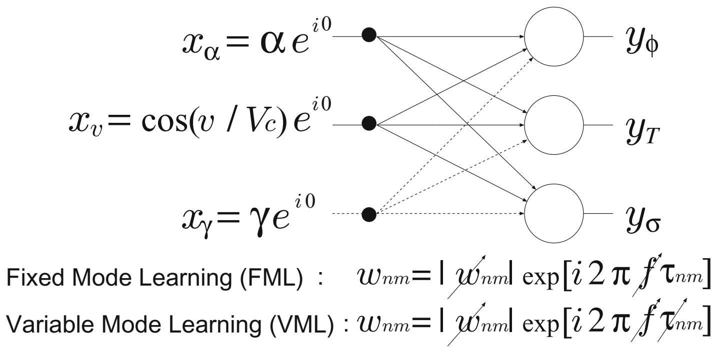


Fig. 10.2 Construction of the neural network. (Reprinted from Fig.2 in [219]: Akira Hirose, Yasufumi Asano, and Toshihiko Hamano: Mode-utilizing developmental learning based on coherent neural networks. In *International Conference on Neural Information Processing (ICONIP) 2004 Calcutta (Lecture Notes in Computer Sciences 3316)*, pages 116–121, Berlin, November 2004, Springer, (C) Springer-Verlag Berlin Heidelberg 2004, with permission.)

information γ works as a sight in the advanced task mentioned below. The sensory signals are real numbers and are fed to the CNN as

$$x_\alpha = \alpha \quad (= \alpha e^{i0}) \tag{10.4}$$
$$x_v = \cos(v/V_c) \quad (= \cos(v/V_c)\ e^{i0}) \tag{10.5}$$
$$x_\gamma = \gamma \quad (= \gamma e^{i0}) \tag{10.6}$$

where angles are represented in radians. Velocity v is normalized by a constant V_c and converted simply into an even function.

On the other hand, the motor signals are obtained at the neural outputs with the constants ϕ_c, T_c and σ_c as

$$\phi = \phi_{\mathrm{c}}\ \mathbf{Im}[y_\phi], \tag{10.7}$$
$$\sigma = \sigma_{\mathrm{c}}\ \mathbf{Im}[y_\sigma] \tag{10.8}$$
$$T = T_{\mathrm{c}}\ \mathbf{Re}[y_T], \tag{10.9}$$

Then, provided that the neural input values and the initial neural connection phase values are chosen at around zero, which means a neutral condition, the bicycle will be controlled almost in neutral by the neural output, i.e., the handlebar is directed straight, the relative human angle of roll is zero, and the wheel torque is moderate. Such a situation may be helpful for the control if the initial weight delays are very small, though they are more at random actually in the experiment. However, note that, this condition is natural and does not violate generality. The values of the constants and the parameters are presented together with other parameters in the next section.

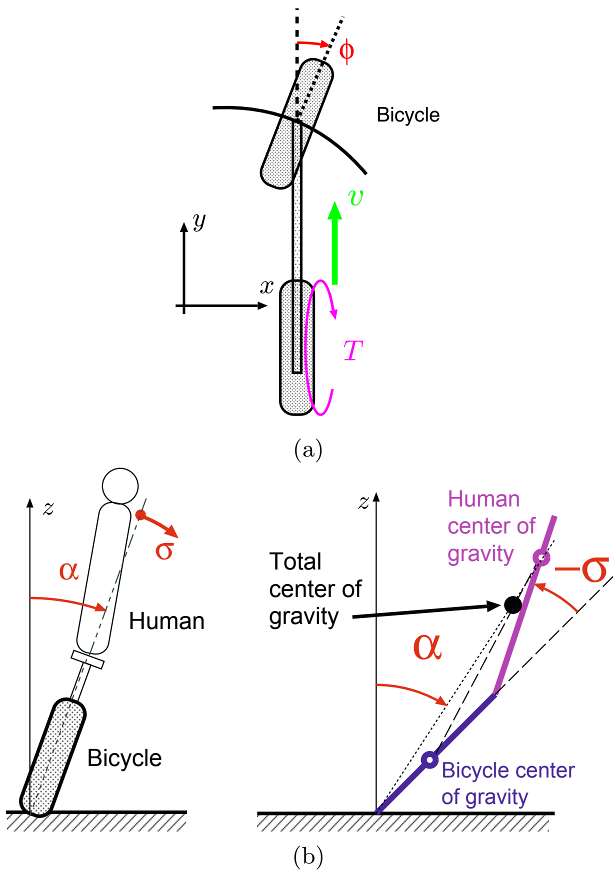


Fig. 10.3 Physical model of the human-bicycle system with variables and parameters: (a)Plan view and rear elevation. (Reprinted from Fig.3 in [219] in figure caption of Fig.10.2 with permission.)

10.3 Developmental Learning in Bicycle Riding

In the present developmental learning experiment, we employ the reinforcement learning having two learning stages. The first one is the random trial where the network changes neural connection weights $|w_{nm}|$ and τ_{nm} at random. The initial values are chosen also at random within certain ranges of the variables, e.g., $|w_{nm}| = 0.01 \sim 0.99$ and $\tau_{nm} = 0.1 \sim 99$ [ms]. We repeat the random trial for certain times, and we find the best trial. This stage is analogous to our rough trials in various ways to ride a bicycle in the real life.

The second stage employs the hill-climbing method by starting at the best condition obtained in the first random trial stage. The hill-climbing process changes the weight components with small fractions $\Delta|w|$ and $\Delta\tau$

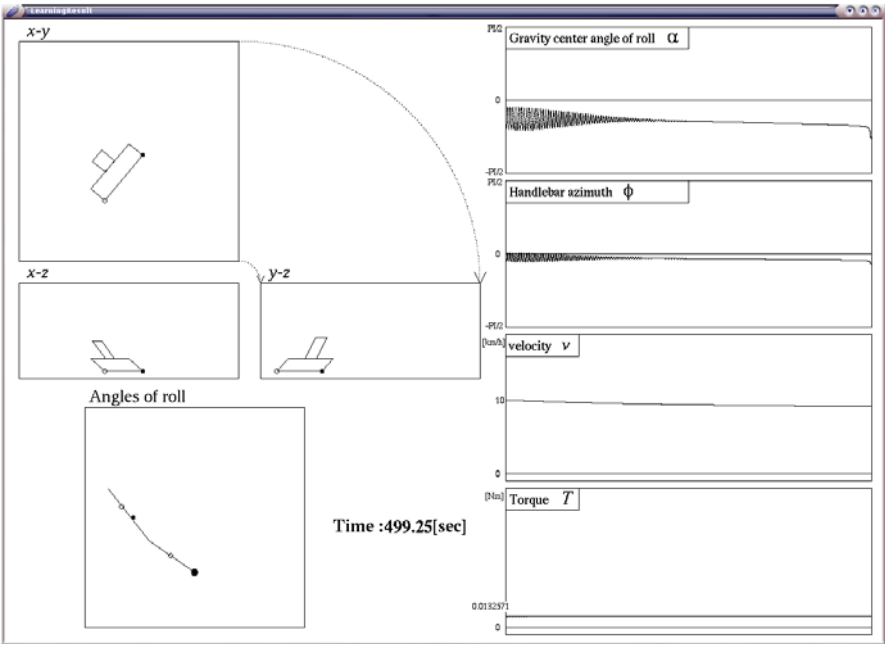


Fig. 10.4 Captured simulator display where the larger rectangle shows a bicycle, while the smaller one shows a human body. (Reprinted from Fig.4 in [219] in figure caption of Fig.10.2 with permission.)

as $|w_{nm}| \longleftarrow \Big| |w_{nm}| + \Delta w \Big|$ (i.e., $|w_{nm}| \geq 0$) and $\tau_{nm} \longleftarrow \tau_{nm} + \Delta\tau$, respectively. If the resulting effect is desirable, the network accepts the small changes. Otherwise, it rejects them. By repeating the process, the network searches a better set of connections. This stage may correspond to learning by iteration of fine adjustment for the human beings.

10.3.1 Task 1: Ride as Long as Possible

First, we try to learn Task 1, i.e., to ride as temporally long as possible. The carrier frequency in (10.3) is fixed at $f_0 = 100$[Hz] so that the behavioral mode is also fixed. The frequency f is kept unchanged at f_0. We call this learning style the fixed-mode learning (FML). In Task 1, the system does not use the direction information γ, which means a *blind* condition.

Figures 10.5 and 10.6 present typical results in Task 1. Figure 10.5(a) shows the riding time before falling down, t_{R}, for every random trial, Fig.10.5(b) shows the riding time t_{R} for the following hill-climbing learning by starting at the best trial condition in the random trial, and Fig.10.5(c) presents the riding locus for the longest-time trial after the hill-climbing

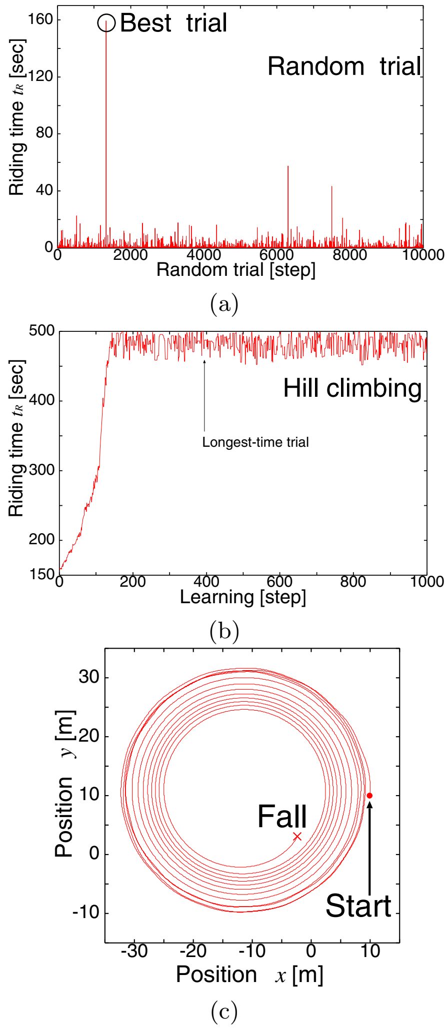


Fig. 10.5 Typical result in Task 1: (a)Riding time t_{R} versus random trial, (b)that versus hill-climbing learning with starting under the best weight-set condition in (a), and (c)riding locus for the longest-time trial after the hill-climbing learning converged. (Reprinted from Fig.5 in [219] in figure caption of Fig.10.2 with permission.)

process converged. The hill-climbing learning is found to extend the t_{R} increasingly, and to accomplish the goal of the long-time riding. However, the locus in Fig.10.5(c) reveals a round course.

However, the obtained behavior is found human-like and very attractive as follows. Figure 10.6 shows the (a)angle-of-roll of the center of gravity α,

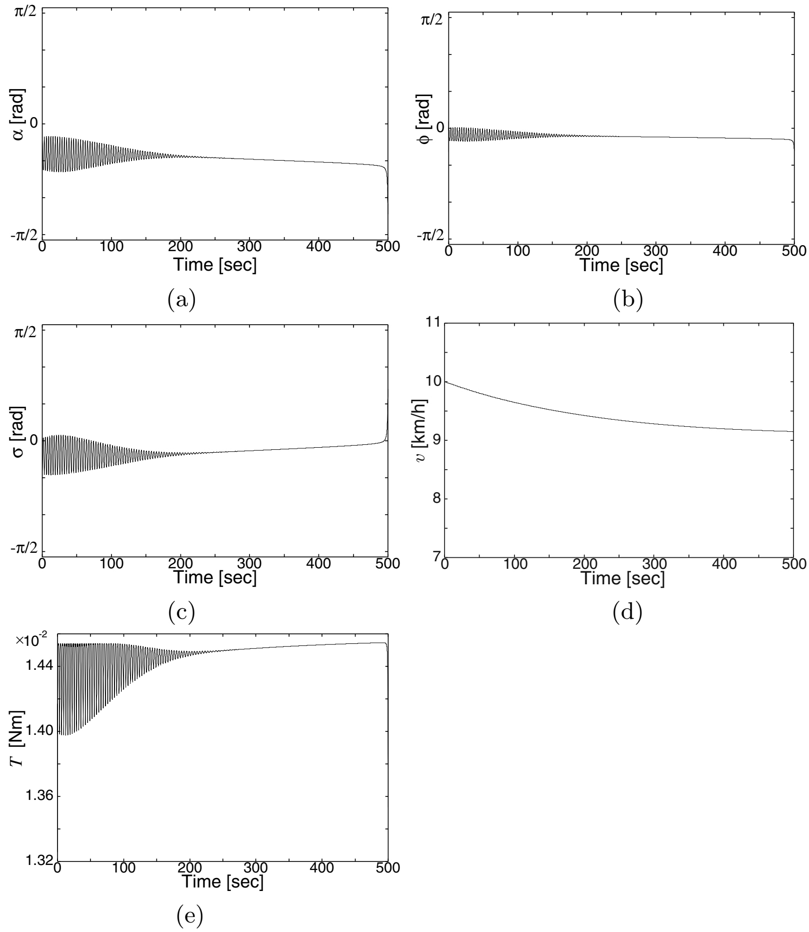


Fig. 10.6 Evolutions of the variables in typical Task 1 result corresponding to the ride in Fig.10.5(c): (a)Angle-of-roll of the center of gravity α, (b) handlebar azimuth ϕ, (c)human-bicycle rolling angle σ, (d)velocity v, and (e)torque T.

(b)handlebar azimuth ϕ, (c)human-bicycle rolling angle σ, (d)velocity v and (e)torque T, all against time, corresponding to the ride in Fig.10.5(c). At the beginning of the ride, the fluctuation of the roll α is large. But gradually the instability disappears. Other variables also present similar evolutions. That is to say, the neural learning has been performed so that a good riding becomes a stable point in the dynamics. This fact is evidence of the appropriateness of the learning.

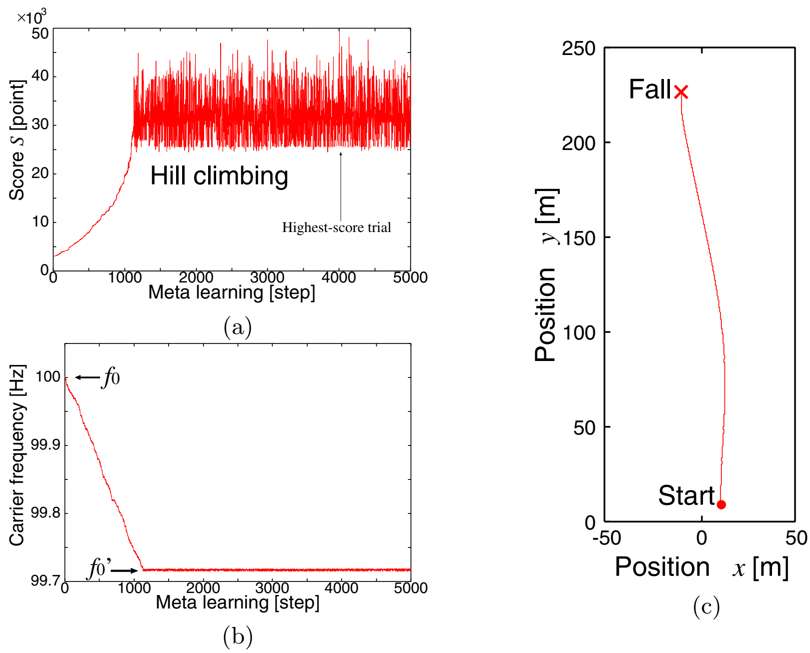


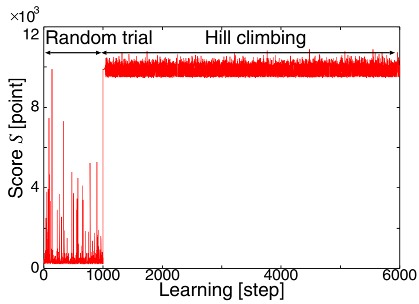
Fig. 10.7 Typical result in developmental VML in Task 2 in hill-climbing process starting with the best result in Task 1: (a)Score S for each riding, (b)self-organization of carrier frequency f, and (c)riding locus for the highest-score riding (Reprinted from Fig.6 in [219] in figure caption of Fig.10.2 with permission.)

10.3.2 Task 2: Ride as Far as Possible

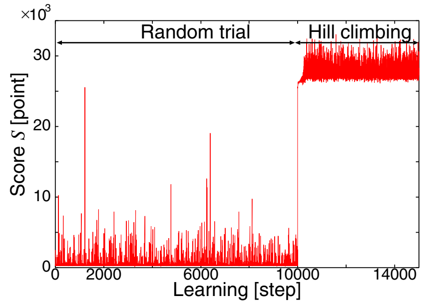
Next, we assign an advanced task (Task 2), i.e., to ride as far as possible. We also prepare an eye to see in which direction the bicycle runs. The direction information $x_\gamma = \gamma$ is fed to the network as mentioned in Section 10.2. This is a *sighted* condition.

We set free the carrier frequency f in (10.3) to enable the network to change the behavioral mode. We call this learning style the variable-mode learning (VML). We expect that the system utilizes the variable frequency. The carrier frequency f is also changed by the hill-climbing method with a frequency fraction Δf as $f \longleftarrow f + \Delta f$ in addition to the hill-climbing learning of $|w_{nm}|$ and τ_{nm}. The frequency shift is equivalent to the variation in behavioral mode. The network searches a mode suitable for a far riding in a self-organizing manner.

In Task 2, we begin with the hill-climbing process by starting at the best result condition in Task 1. In addition, we define an evaluation function (score) S of the far riding so that, the further the bicycle runs, the higher the score becomes. (See, e.g., Ref.[219], [220] for details.)



(a)



(b)

Fig. 10.8 Typical examples of scores in direct FML in Task 2 versus learning steps with (a)1,000- or (b)10,000-times random trials and following hill-climbing learning. (Reprinted from Fig.7 in [219] in figure caption of Fig.10.2 with permission.)

Figure 10.7 shows a typical result of the developmental VML in Task 2. The starting condition is the best result one in Task 1. In Fig.10.7(a), we find a quick increase in the score S. Figure 10.7(b) presents the variation of the carrier frequency f, which works as the behavioral mode parameter. It moves self-organizingly from $f_0 = 100$[Hz] to an optimal value f_0'. That is, the network finds out the mode most suitable for the environment by itself.

Figure 10.7(c) shows the riding locus of the highest-score result after the hill-climbing learning converged. In comparison with 10.5(c), the course has been clearly straightened. The result is obtained quickly by the mode modulation of the longest-time-ride condition to adapt to the new environment, i.e., the advanced task of long-distance riding.

10.3.3 Comparative Experiment: Direct FML in Task 2

We also conduct experiments on developmental FML, direct VML, and direct FML. (See details in Ref.[220].) Figure 10.8 shows typical results of the direct FML experiments without the learning in Task 1. We repeat random trials for 1,000 or 10,000 times and, then, move to hill-climbing learning afterward. The direction information γ is fed to the neural network. The initial state is statistically the same as that of Task 1.

In Fig.10.8(a) and (b), we find that, in the random trial, a high-score probability is very low. Moreover, even in the hill-climbing learning, the score increases only slightly, which suggests that the random trial does not bring the network to the vicinity of a truly ideal state.

10.3.4 Comparison between the Results

When we compare the developmental VML (Section 10.3.2), developmental FML, direct VML, and direct FML (Section 10.3.3), we find that the developmental VML is the most effective method in total [220]. In Fig.10.7(b), we also find that the carrier frequency self organizes to realize such effective learning. We can see that the network learns similar or advanced tasks quickly by changing the internal mode parameter.

10.4 Summary

We have presented the idea of the mode-utilizing developmental learning architecture based on the coherent neural network. The network learns similar or advanced tasks incrementally by using its cumulative skill by changing the behavioral mode-tuning parameter, i.e., the carrier frequency of the coherent network. The mode parameter has been found adjusted self organizingly and smoothly in the developmental learning. The developmental learning and required architecture will be of growing importance in building so-called brain-like systems.