Learning Multiple Goal-Directed Actions through Self-Organization of

a Dynamic Neural Network Model:

A Humanoid Robot Experiment

Ryu Nishimoto, Jun Namikawa and Jun Tani

Brain Science Institute, RIKEN

2-1 Hirosawa, Wako-shi, Saitama, 351-0198 Japan

Tel +81-48-467-6467, FAX +81-48-467-7248

E-mail tani@brain.riken.go.jp

Original Contribution

Running Title

Learning of Multiple Goal-Directed Behaviors

Key Words

Learning, Actions, Initial sensitivity, CTRNN, Self-Organization, Humanoid

Abstract

The paper introduces a model that accounts for cognitive mechanisms of learning and generating multiple goal-directed actions. The model employs a novel idea of so-called the "sensory forward model" which is assumed to function in inferior parietal cortex for generation of skilled behaviors in humans and monkeys. A set of different goal-directed actions can be generated by the sensory forward model by utilizing the initial sensitivity characteristics of its acquired forward dynamics. The analyses on our robotics experiments show qualitatively that (1) how generalization in learning can be achieved for situational variances, (2) how the top-down intention toward a specific goal state can reconcile with the bottom-up sensation from the reality.

1 Introduction

How can robots learn to generate actions flexibly in achieving given goals while giving off fluency of acquired skills, as we humans do, even when they are exposed with situational changes in the environment? Upon approaching this problem, there are several issues to be contemplated.

First, when robots face variability in each situation, generalization capability in learning and generating behaviors among different situations become the essential problems. If there is no variability, the problems of the behavior generation turn out to be trivial ones of just remembering and regenerating a fixed trajectory to the goal. Bianco and Nolfi (Bianco & Nolfi, 2004) demonstrated the development of such generalization capability by robots in grasping objects in simulation. In their study, the simulated robot becomes able to grasp objects located in arbitrary position and orientation in an adopted workspace because the robot is evolved through trials under such variable situations. The current study investigates how this type of generalization in object manipulation by real robots can be achieved through supervised training of them.

Second, the way of interactions between the top-down intentions toward goals and the bottom-up sensation of the real world become crucial in achieving goal-directed action in a stable manner(Tani, 2003). Given a goal, the robot could have an anticipatory image of the future sensory-motor flow that leads to the goal. However, such a top-down image should be modified adaptively by responding to the current sensory reality by having dense interactions. In order to organize such dense interactions between the top-down intentions and the sensory-motor reality, the both sides should share a direct and smooth interface between them. The current paper argues that such interface should be best built on dynamical systems defined

in continuous time and space domain.

Third, we assume that cognitive agents should be able to attain representations of skilled behaviors such that the same representation can be utilized both in generation and recognition of the behaviors. More specifically, if the sensory flow of an own act can be imaged by the top-down anticipation, an identical act by others could be accounted by utilizing the same internal model that generates the image (Oztop & Arbib, 2002; Oztop, Wolpert, & Kawato, 2005; Tani & Ito, 2003; Ito & Tani, 2004). This view has been inspired by biological observation of mirror neurons (Rizzolatti, Fadiga, Galless, & Fogassi, 1996) in monkeys, as will be described shortly after.

Forth, the context-dependency in generating action sequences toward goals is also an indispensable nature, of which aspect was metaphorically depicted as "kinetic melody" by Luria (1973). For example, let us consider situations that we grasp a cup. Then, we notice that the exact motor trajectories of grasping a cup should be affected by next motor act to follow them as well as the entire goals of whether to drink a cup of water or to clean off the cup. The claim here is that the context-dependency involves not only with the information processing in brains but also with the physical body movements. The whole motor behaviors should be generated fluently by capturing the contextual flow of intended goal-directed actions.

Now, let us look at some of neuroscience literatures to explore biological observations corresponding to our interests. Neuropsychological studies investigating various apraxia cases, including ideomotor apraxia and ideational apraxia (Liepmann, 1920; Heilman, 1973), have suggested that inferior parietal lobe (IPL) should be an essential site to represent a class of behavior skills, especially related to object manipulation and tool usage. Furthermore,

electrophysiological studies of monkeys by Fogassi et al. (2005) recently showed that IPL functions both for generating and recognizing the goal-directed behaviors by having dense interactions with cells in ventral premotor (PMv) which are known as mirror neurons (Rizzolatti et al., 1996). Based on these evidences, the current paper assumes that IPL might function as so-called the "sensory forward models" by having certain anticipatory mechanisms for coming sensory inputs in achieving a specific goal-directed action. On this account, it is furthermore assumed that the goal of the action might be set in PMv which is sent to the sensory forward model assumed in IPL as the inputs.

The current paper introduces a dynamic neural network model which attempts to model IPL-PMv interactions in learning, generating and recognizing a set of goal-directed skilled behaviors. Our model is different from the one using local representation scheme proposed by Demiris and Hayes (2002) in which each goal-directed behavior is learned in a different forward model. Our model employs a distributed representation scheme because the scheme can afford generalization by learning different goal-directed behaviors in a single network (Tani, Ito, & Sugita, 2004). One central assumption in our modeling is that trajectories of different goal-directed behaviors can be generated by means of the initial sensitivity characteristics of a single forward dynamics model assumed in IPL. And PMv is assumed to set a different initial state for the IPL forward dynamics model corresponding to each goal-directed behavior. On the other hand, observed behavior patterns can be recognized by inferring the corresponding initial states inversely through the forward model.

Our proposed model is implemented in a humanoid robot experimental platform for a task of learning simple object manipulations with multiple goals using a simplified vision. Our experiments will clarify (1) what dynamic structures can be self-organized for learning

multiple goal-directed behaviors with utilizing the initial sensitivity characteristics of the employed dynamic neural network model, (2) how generalization in learning can be achieved through experiencing sensory variances introduced in the experiment. (3) how the top-down intention toward a specific goal state embedded in the initial state can reconcile with the bottom-up sensation that brings fluctuates in the outer world reality.

The next section introduces our abstract modeling of IPL-PMv networks followed by its implementation into a dynamic neural network model, the results and analyses of the robotics experiments, and final discussions.

2 Modeling IPL-PMv networks

In the current paper, the generation and recognition pathways of actions assumed in the IPL-PMv networks of human brains are modeled as follows.

In generating goal-directed actions, PMv sets the initial state of a forward model, which is assumed to exist in IPL, as corresponding to the current goal in the preparatory period of the actual action. This forward model is slightly different from the conventional forward model (Kawato, Furukawa, & Suzuki, 1987; Wolpert & Kawato, 1998) assumed in cerebellum. The conventional forward model predicts the next time step sensory outcomes regarding to the current action taken and the current sensory state given. The proposed forward model, however, predicts the next time step sensory state solely with the current sensory state. We may call it as the sensory forward model. Given with an initial state representing a specific goal information, the sensory forward model generates both prediction of the visual inputs flow and the proprioceptive one reaching to the goal state. The prediction of the visual inputs is sent to the early vision area including V1 and that of the proprioception does to

6

the primary sensor area S1. It is noted that the predicted changes of proprioceptive state include that of the posture. Therefore our wild assumption here is that this information might be sent further to cerebellum and/or M1 for the purpose of generating corresponding motor commands to actualize the predicted postural changes. The analogy to robotics is that IPL generates sequences of target encoder values for all joints and then cerebellum and M1 as motor controllers generate the corresponding motor torque sequences. (See Figure 1(a) for an illustration of the generation pathway.)

Nextly, we explain that goals of behaviors by others can be recognized by going inversely through the generation pathway described above. For example, in case of recognizing a person to manipulate an object, the sensory forward model can predict how the visual image, the extracted features about positional relations between the object and the hand evolves in time if the initial state representing the goal of the person is correctly set. (Although the coordinate transformation from the egocentric view to the object center one is another essential problem, the current model does not cover this problem. The problem is left for future studies.) If the initial state is set incorrectly in the sensory forward model, the prediction for the coming visual sensory image generates error. In this situation, the prediction error can be minimized by adapting the initial state values to optimal ones. If the optimal value is eventually found, it can be said that the goal of the behavior by this person is recognized. (See Figure 1(b) for an illustration of the recognition pathway.)

Previously, Ito and Tani (2004) and Oztop et al. (2005) showed a different type of forward model in which the goal states are represented by a constant input vector to the forward model. Our previous study (Nishimoto & Tani, 2004) suggested that the current scheme utilizing the initial sensitivity characteristics can be extended to learning of more complex

trajectories of having tree-like successive branching structures.

3   Implementation to CTRNN

The sensory forward model assumed in IPL is built by using a continuous-time recurrent neural network (CTRNN) (Williams & Zipser, 1989; Doya & Yoshizawa, 1989; Beer, 1995). CTRNN is employed because its neuronal activation dynamics defined in continuous time and space domain can access directly to analog sensory signals of the robot without supposing arbitrary interfaces.

Another reason for this employment is that CTRNN is expected to capture the context-dependent properties hidden in adopted task behaviors by utilizing its internal units, so-called the context units. The studies on conventional RNN models (Elman, 1990; Jordan, 1986) have shown such effects of having the context units in generating sequences. In the case of CTRNN, its context loop can afford long time correlations in its time development by adequately tuning the time constant parameter of the neuronal activation dynamics. Our trial in the current study is to utilize the initial sensitivity characteristics of this context state dynamics for representing different goal-directed behaviors by a single CTRNN. It is noted that the CTRNN architecture does not correspond directly to the real anatomical neuronal structures in IPL. In the current study, the CTRNN is used in the connectionist level abstraction for the purpose of attaining the forward model functions. The usage of the error back propagation learning scheme (Rumelhart, Hinton, & Williams, 1986) which may not represent real neuron synaptic modulation mechanisms can be accounted by the same reason.

The CTRNN is trained with multiple sensory (visuo-proprioceptive) sequences acquired

during the teaching phase of the robot in a parallel manner with off-line operation. Those sensory sequences include the ones for teaching different categories of goal-directed behaviors with having certain environmental variations. Then, the goal of training the CTRNN with these teaching sequences is to attain the sensory forward model that can regenerate these sequences with certain generalization. On this purpose, the training process of the CTRNN determines two things, namely synaptic weights common to all sequences and the initial context state (the context unit activation values in the initial step) specific to each sequence. It is noted that no categorical representation for goals are given explicitly to the network. It, rather, appear as the results of self-organization through both adaptation of synaptic weights and the initial context state (Nishimoto & Tani, 2004).

The CTRNN receives visuo-proprioceptive sequence patterns represented as (s_t, m_t). Here, s_t represents vision-based sensory information at time step t, representing direction of the object attended by the robot video camera. The other component of the sensory information m_t represents proprioception of the arms of the robot by reading encoder values in the arm joints. Essentially, the CTRNN receives the inputs consisting of different modalities of sensation and mingles those inputs together to generate predictions of their time developments in future.

We designed our CTRNN model, shown in Figure 2, by modifying the Jordan-type RNN–a discrete time model. Figure 2(a) shows the training mode of the CTRNN. Figure 2(b) shows the open-loop generation mode when the robot actually acts on the environment.

Before going to detailed mathematical descriptions of the model, we explain the basic mechanism in a intuitive manner. As shown in Figure 2, the network consists of three layers, namely the bottom layer, the hidden layer and the top layer where there are reentrance loops

9

from the top layer to the bottom layer. The bottom layer consists of groups of units in which the visible units a^X corresponds to the visuo-proprioceptive state (s, m) and the context units a^C does to the internal state of the network. The activation states of these units are propagated into the hidden units in the hidden layer through the synaptic weights w_{ij}^H and further to groups of the output units o^X and o^C in the top layer. The output units here are utilized in a different way from usual perceptron-type networks or RNNs. o^X and o^C mostly represent δ changes of a^X and a^C. Actually, the time development of a^X and a^C in the bottom layer are obtained by integrating o^X and o^C in time through the reentrance loop. Next, the detailed computation scheme for the CTRNN are described.

Forward dynamics

The visuo-proprioceptive sequences are predicted by means of the forward dynamics of the network. First, the hidden layer activation a_t^H is computed in the forward dynamics. The potential of the ith hidden unit $u_{i,t}^H$ is obtained by summing the activation propagated from the bottom layer through the synaptic weights w_{ij}^H. Then its activation value is computed by applying the sigmoid function to the potential. This is described as:

$$u_{i,t}^H = \sum w_{ij}^H a_{j,t}^{X,C} + b_i^H \tag{1}$$

$$a_{i,t}^H = sigmoid(u_{i,t}^H) \tag{2}$$

where b_i^H represents the bias value for the ith hidden unit. Next, the top layer activations of o_t^X and o_t^C are computed in the same way using hidden layer activation values.

$$o_{i,t}^{X,C} = sigmoid(\sum w_{ij}^O a_{j,t}^H + b_i^O) \tag{3}$$

10

Then, time development of the potential of the bottom layer units, u^X and u^C, are obtained by integrating o^X and o^C by following the first order differential equation of Eq.4.

$$\tau_{X,C}\dot{u}_i^{X,C} = -u_i^{X,C} + (o_i^{X,C} - 0.5)/\alpha \tag{4}$$

where $\tau_{X,C}$ and α denote a time constant and amplification coefficient, respectively. The actual update of these values are computed with the numerical approximation of Eq.5.

$$u_{i,t+1}^{X,C} = -u_{i,t}^{X,C}/\tau_{X,C} + (o_{i,t}^{X,C} - 0.5)/(\alpha \cdot \tau_{X,C}) + u_{i,t}^{X,C} \tag{5}$$

Finally, the next time step activation of the visible and the context units are obtained by taking the sigmoidal output of the potential.

$$a_{i,t+1}^{X,C} = sigmoid(u_{i,t+1}^{X,C}) \tag{6}$$

Here, $a_{i,t+1}^{X}$ represents the closed-loop forward prediction for visuo-proprioceptive state at the next time step (s_{t+1}, m_{t+1}).

The open-loop one-step prediction can be performed with utilizing the inputs from the external rather than the self-predicted ones as shown in Figure 2(b). In the open-loop mode $u_{i,t}^{X}$ in the right-hand side terms in Eq.5 is replaced with $\bar{u}_{i,t}^{X}$ which is the potential value representation of the actual visuo-proprioceptive state. Here, $\bar{u}_{i,t}^{X}$ can be obtained by taking the inverse sigmoid of $\bar{a}_{i,t}^{X}$. The open-loop prediction generates actual movements of the robot using the actual sensory inputs. On the other hand, the closed-loop prediction can generates lookahead future prediction without accessing the actual sensation by which "mental simulation" of the robot can be achieved (Tani, 1996).

It is noted that the time constant $\tau_{X,C}$ determines the dynamic characteristics of the network substantially. If the $\tau_{X,C}$ is set larger, the network easily adapts to slower dynamics

of the sensory flow. Otherwise it does to faster one. The time constant for the context

units τ_C should be set larger than the one for the sensory units τ_X because the context

units are required to preserve long range time-correlations related to goals while the sensory

units are expected to respond to the environmental situations promptly. The prediction of

the proprioception m_t is utilized to generate the actual motor command of the robot by a

position feedback controller at each joint. On the other hand, the prediction of the vision

state s_t is used for recognition of the observed behavior, as will be described later.

Training

The goal of training the network is to minimize the error between the teaching sequence

pattern $\hat{a}_t^X$ given from the outside and the predicted sequence pattern a_t^X of own generation.

The network learning method uses the general Back Propagation Through Time (BPTT)

algorithm (Rumelhart et al., 1986). Using the BPTT algorithm, the network obtains shared

weight connectivities for all the given teaching sequence patterns. At the same time, in the

proposed model, the initial context state a_0^C is adapted corresponding to each of different

goal-directed behaviors of trained. It is aimed that each desired goal-directed behavior can

be generated by setting its corresponding adapted value into the context units at the initial

step after the training.

In the actual learning process, an update of a connective weight from the i-th unit to the

j-th unit at the nth learning iteration step are obtained by using the delta error δ_i (Rumelhart

et al., 1986) back-propagated to the i-th unit in the following.

$$\Delta w_{ij}(n+1) = \eta \delta_i a_j + \alpha \Delta w_{ij}(n) \tag{7}$$

This training of the connective weights are conducted in a parallel manner for all the teaching

sequences. This means that at each epoch of training the forward activation and the BPTT are conducted for all sequences once and then the connective weights are updated using the accumulated delta error. In addition, an update of the initial context state corresponding to each training sequence pattern is obtained simultaneously by utilizing the delta error δ_i back-propagated through time steps to the context unit at the initial step.

$$\Delta a_{0,i}^C(n+1) = \eta \delta_{0,i} + \alpha \Delta a_{0,i}^c(n) \tag{8}$$

In the actual experiment, only two dimensions of the initial context state are adapted and all other dimensions are set as 0.5 for the sake of easiness in analysis described later. Our preliminary examinations showed that the learning can be achieved by adapting any number of dimensions in the initial context state if they are more than one.

Generation and recognition

After the network is trained, the network can be operated both in the generation mode and the recognition mode. In the behavior generation mode, once values for the initial context state (a_0^C) are set, the network generates the corresponding visuo-proprioceptive sequences with the open-loop. In the behavior recognition mode, the vision state sequences $\hat{s}_t$ are given to the network as a target sequence to mimic. Then, optimal values for the initial context state is searched by following Eq.8 such that the error between the sequence pattern generated from this initial context state in the closed-loop mode and the target sequence pattern can be minimized. It is noted that the optimization of the initial context state by the error back-propagation in the recognition mode is just one possible implementation in the connectionist approach. The same function in terms of error regression could be realized by other neuronal mechanisms in real brains.

13

4 Experiment setup

The experiment is aimed at autonomously generating three types of operational actions on objects, as shown in Figure 3, based on learning. In those goal-directed actions, both arms of the robot start to move from a fixed home position in order to approach and then to grasp an object placed in front of the robot by both hands. And then, the object is either (a) lift up, (b) moved to right, or (c) moved to left. The time constant for the context units τ_C should be set larger than the one for the sensory units τ_X because the context units are required to preserve long range time-correlations related to goals while the sensory units are expected to respond to the environmental situations promptly.

The experiments are conducted using a small-scale humanoid robot produced by SONY (see Figure 3). The joints below the waist of this robot are fixed to a platform, and the tasks only involve movement of the head and arms of the robot. The robot has two arms each of which has 4 degrees of freedom without fingers. The arm stretch is about 30 cm. The encoder values of these arm joints are received as the current proprioceptive inputs m_t and its next step prediction m_{t+1} is sent to the joint PID controller to generate their actual movements. The time sampling of the sensory-motor state of the robot is 150msec determined by the hardware specification. This is equal to the numerial integration step interval of the CTRNN.

A workbench is set up in front of the robot and a cubic object approximately 6 cm3 volume is placed on the workbench to serve as the goal object. This object is red-marked so that the robot can always recognize the object's position using its vision-tracking program. In order to examine the generalization capability of the object grasping action regarding to variances in object position sensed by vision, the robot is trained to act on the object initially

located among three different positions, "right", "left" and "center" in front of the robot. The position difference in the object center between the right position and the left one is 10 cm.

The head of the robot contains a built-in color vision camera that is equipped with a program that moves the neck of the robot to maintain the red-marked object constantly in the center of the field of vision. This vision tracking program always follows the object, regardless of the action output from the network. The direction of the robot head in terms of encoder values of two neck joints represents the object position relative to the robot. This 2-dimensional vector corresponds to the visual input denoted as s_t in the previous section.

In the training, the experimenters guide the robot arm movements by grasping the arms in order to generate the desired trajectories. The experimenters guide to move the robot arms smoothly as much as possible in achieving the goals of actions. The encoder values as well as the visual inputs (2-dimensional head direction) are recorded during the guidance for their later usages for the off-line training of the CTRNN. For the "lift up" action, the arms are guided such that the object is lift up for 10 cm. For the actions of "move left" and "move right", the arms are guided such that the object is moved to extreme left and right within the mechanical limitations of the robot. The arm trajectories are guided for three types of the goal-directed behaviors with the three different object positions for each. Therefore, there are 9 categories of the teaching sensory sequences where each category sequence is repeated 3 times. As the total, 27 sensory sequences are sampled during the supervised teaching phase. The CTRNN is trained with these teaching sequences in a parallel manner for 50000 epochs.

The action generation tests are conducted in the following two conditions, (1) with locating the object at arbitrary positions between the left position and right one defined in the

training phase and also in the outside of this trained region, (2) with changing the object position from "center" to "left" or "right" abruptly at arbitrary time step after the onset of the robot movement. Here, the "lift up" action is defined as successful if the robot lifts the object at least 5 cm above the floor. The "move left" and "move right" actions are defined as successful if the robot moves the object at least 5 cm from the original position.

The CTRNN used for this experiment is allocated with 20 hidden units, 20 context units and 10 sensory inputs (2-dimensional vision-based inputs and 8 -dimensional proprioceptive inputs for 8 joints in two arms). The time constant for the context loop dynamics τ_C is set much larger than that for the sensory loop τ_X because the context loop dynamics is aimed to capture long time correlation structures hidden in the training patterns. In the following experiment τ_C and τ_X are set as 20.0 and 7.0, correspondingly.

5   Results

The repeated training examinations with randomly set different initial weights revealed that the mean square error (MSE: average square error per sensory unit per step over all teaching sequences) converges around 0.0003. By taking one case of the trained weights with its MSE converging to 0.00028, the robot performances of generating the goal-directed behaviors were examined.

5.1   Action generation

Each goal-directed action was generated with setting the initial context state with the value that had been adapted in learning the same goal-directed action with the object located in the center. The experiment results showed that the robot generates each goal-directed

behavior successfully when the object is positioned in the center. Figure 4 shows the time-developments of the head direction representing the relative position of the object and the context unit activations for each behavior. It is shown in the left-hand side of Figure 4 that the head direction starts to change largely after around 30 steps with developing a different profile depending on each goal. The right-hand side in this figure shows that some context units activations develop differently for each goal and some of them do similarly. Especially, two bold lines representing the activation of the context units, of which initial values are adapted, develop in different ways substantially with starting from different initial values depending on the goals. It is noted that these two context units activation states develop differently even when the same movement trajectory is generated in approaching to the object in these three actions. It is speculated that the context state work as dynamic memories to initiate each different goal-directed movement that follows the same movement of approaching to the object in the initial period.

5.2 Generalization

Additionally, we examined the generalization capability of the learned network. The generalization for the object position shifts is expected for the width direction but not for the depth direction because the learning was conducted only for the former case. Therefore, action generation was tested with varying the object position from left to right within the same range used in the teaching phase. The initial context state for generating each goal-directed behavior is set with a representative value, which had been obtained for the case of the object positioned in the center in the learning phase, for all trials with different object positions. The experiment results showed that each goal-directed behavior is generated successfully re-

gardless of the object position. This means that the initial context states are adapted in the learning such that they are sensitive for the goal categories but not for the object position in behavior generation. Starting from one representative initial state for each goal, the CTRNN dynamics successfully generates the corresponding goal-directed motor trajectory with adapting to the position differences of the object.

We also examined how much this generalization capability of the robot can tolerate with the object position variances beyond the range of trained. For this purpose, the "lift up" action was tested with varying the object position from the left position of trained to 5cm further left with each 1cm interval. The initial context state was set with the representative value for the "lift up" action as it was done in the previous experiment. The success rate for each position with 5 times trials is shown in Figure 5. It is shown that the trials are successful with 100 percentage when the object is positioned within 2cm further left from the trained left position. The robot starts to fail with the object positioned at 3cm further left. With more than 4cm further left the robot fails completely. In these failure cases the robot hands cannot approach to exact positions required to grasp the object. These results show that the generalization capability for the object position variance is extended to the outside of the trained range to some extent.

5.3 Real time adaptation

Finally, we tested how much the robot can adapt to the sudden changes in the environmental situation in a real time manner. For this purpose, we tested the situation that the object is moved suddenly either from the current left position to the right position or vice versa after the robot starts to generate the "move up" action. The timing of the interference was

changed between 0th step and 20th step from the onset of the robot movement. Note that the robot normally touches the object around 30th step. The success rate for each timing with 10 times trials for each is shown in Figure 6. It is seen that the action failure starts with the interference at 10th step after the onset of the robot movement. The success rate becomes zero at 18th step after the onset. The result indicates that the robot acquired a real-time adaptation capability to some extent through the learning.

5.4   Analysis on the initial context state map

We further examined how the set of trained goal-directed actions are internally represented by constructing so-called the initial context state map. Figure 7 plots the 2-dimensional initial context state adapted for each training sequence. In Figure 7(a) the plots are made with using category labels representing different goals. Figure 7(b) does with different object positions with sharing the same data which is aimed to show possible effects of the position variances to the internal representation of the goals. From these two plots it is seen that a topological map is self-organized between the initial state and the goal states but not for the object positions. We see categorical clusters in the initial state map as corresponding to different goal states but not for the object positions. This analytical result shows that the initial context state map is self-organized as sensitive to the goals but not for positions of the object. This implies that the adaptability in the motor behavior of approaching to the object in various positions is achieved by utilizing the sensory feedback. On the other hand, the goal information is well preserved in the context-loop dynamics as a short-term memory without being affected by the position variances. The above experimental results conclude that the network has achieved both properties of the adaptability to the environmental variances and

the robustness in achieving the goals.

5.5   Recognition of actions by others

We examined the behavior recognition processes of the robot when the object is moved by the experimenter in the same way as the robot does. As it has been described in the previous section, the recognition is a process to search for an optimal 2-dimensional vector of the initial context state that best fits with the observed behavior. Figure 8 shows how the value of the initial context state is searched from its neutral value $(0.5, 0.5)$ during the recognition process of the visual input sequence pattern s_t for each observed goal-directed behavior. It is seen that each value of the initial context state (denoted by trajectory reaching to black category label) converges to neighbor of the appropriate goal cluster which was shown in Figure 7. It can be said that the network recognizes the observed goal-directed actions correctly using the proposed scheme.

5.6   Analysis on effects of time constant changes

It is noted that the success of the robot actions largely depends on setting of the time constant parameters τ_C and τ_X of the CTRNN. In the all experiments shown up to here τ_C and τ_X are set as 20.0 and 7.0, correspondingly. This subsection shows how the changes of the time constants affects the ways of the CTRNN self-organization and that of behavior generation.

For this purpose, the CTRNN was trained with the original teaching data for two cases, the one with smaller time constant ($\tau_C = 5.0$, $\tau_X = 1.75$) and the other with larger one ($\tau_C = 40.0$, $\tau_X = 14.0$). Although the detailed learning results differed according to the time constants after 50,000 steps of training, there was no dispersion for any of the experiment

conditions, and virtually all the time sequence patterns could be learned. Using each of the CTRNNs after training, we conducted the action generation test for each goal with varying the object position in the same way as it had been conducted in the previous experiment. The initial context state is set with the value that is adapted in the training of each goal-directed action with the object located in the center position.

The following summarizes the results briefly. In the case with smaller time constants, the robot cannot achieve the goals properly. It was, however, found that the adaptability to the object position variances can be much improved in this case i.e. – the robot, at the least, can grasp the object regardless of its position but cannot continue the successive goal-directed behaviors of specified. This is because the goal information preserved in the context state dynamics is likely to be lost due to its small time constant during the motor act of approaching to the object as affected by its position variances.

The robot cannot achieve any goals again in the case with larger τ parameters. In this case although the context state profiles tend to preserve enough differences among different goal-directed actions during movements of the robot, behavior generation cannot adapt well to changes of the object positions because the motor behaviors become too sloppy due to large τ parameters.

Figure 9 shows the initial context state map as plotted for the smaller τ case and the larger τ case for their comparisons. They are plotted twice with using different category labels representing the goals and with those for the object positions in the same way with the previous plots. (Those plots should be compared also with the plots made with the original τ parameters shown in Figure 7.)

In the case of the smaller τ parameters shown in Figure 9 (a), it is seen that the initial

context state is not categorized with the goals but it does with the object positions. On the other hand, in the larger τ case shown in Figure 9 (b), categorizations with the goals can be seen. Furthermore, sub-categorization with object position can be seen at each goal category.

These analysis results indicate how the dynamical structures can be self-organized with utilizing the initial sensitivity characteristics depending on the time constant parameter in the CTRNN. With having smaller time constants, the initial context state map cannot be self-organized as sensitive to differences of goal states because the dynamics cannot afford long time correlations from the initial state to the goal state. However, it does as sensitive for differences for the object positions because the dynamics can afford time correlations at the least for the shorter period until the state of grasping. With having large time constants, the initial context state map is self-organized with forming a hierarchical categorical structure, namely main categories for end-point goal states and sub-categories for middle-point grasping state. Because not only the goal but also the object positions for grasping become sensitive to the initial state, the dynamics with large time constant cannot adapt to the object position changes as on-line by deviating from the already determined trajectories by the initial state set in the beginning. Finally with setting the time constants with the original values, which are between the two extremes, two different functions of goal-directed one and environment sensitive one are achieved as polymerized in one self-organized dynamical structure. The former is achieved by means of the top-down-like initial sensitivity dynamics and the latter by the bottom-up-like self-adapting dynamics utilizing the sensory feedback.

22

6 Discussion

An interesting point in the current experiment result is that the goal-directed actions are generated by dynamically chunking sensory-motor flow into a set of action primitives. The robot action starts from "approach to object" and then it branches to either of "lift up", "move to left" or "move to right" based on the initial state of the context units. Pitchfork bifurcation (Wiggins, 1990) can explain this mechanism of branching in the proposed CTRNN scheme (see an illustration shown in Figure 10.) It is intuitively explained that there is only one stable fixed point in the projection of x which encodes action primitive A before the branching point. Then after the branching, it bifurcates to two stable fixed points that encode action primitive B and action primitive C. (Note that the current robot experiment case corresponds to trification rather than bifurcation. We employed this bifurcation example for its simplicity.) It is essential that each action primitive is encoded by the convergent dynamics because the convergent vector flow provides inherent robustness in generating motor pattern. Now, let us consider how this dynamic mechanism can account for adaptability to situational changes in environment. For example, back to the current robot experiment case, if the position of the object is changed gradually, the stable fixed point of action primitive A, actually encoding "approach to object", would shift accordingly. This shift of the stable fixed point should be responsible for the adaptability of robots to their situational changes.

Pitchfork bifurcation provides a mean to shift motor patterns drastically by dynamically switching from one action primitive to another. The way of the switching depends on the goal information embedded in the initial state of the context units in the CTRNN. Different time developments of context units dynamics starting from different initial states result in different motor act sequences. Nishimoto and Tani (2004) showed that a conventional RNN

of Jordan-type (Jordan, 1986) can learn complex sequences of having tree-like branching structures by utilizing the same initial context state adaptation scheme in their simulation study. However, this scheme of using the RNN with discrete-time mapping function has difficulty in learning smoothly changing long time steps pattern profiles. On the contrary, the CTRNN which is orginally defined in continuous time and space domain can learn such profiles with long time correlation provided that its time constant is set adequately, as have been shown in our experiments.

The scheme proposed in the current paper provides one possible mechanism to achieve compositionality in which action primitives are adequately combined into desired goal-directed actions. The idea of decomposition of whole actions into sequences of reusable action primitives are originally considered by Arbib (1981) as the motor schemata theory. However, when we think about the action primitives, they should not be regarded as concrete objects as like a set of template patterns stored in a computer database. Instead, each action primitive should be acquired as enough "elastic" such that it can be utilized flexibly in various situations. When action primitives are combined sequentially into a goal-directed action, those behavior primitives are not independent elements of the whole. Instead, they are bound together through mutual fitting with sharing the same stream of the context flow heading toward the specified goal. Therefore, their sequences can be generated fluently and seamlessly of which appearances could be well expressed by a metaphor of "kinetic melody" (Luria, 1973). This essential feature originates from the intrinsic properties of the CTRNN model which is defined in the continuous time and space domain. The CTRNN forward dynamics associated with large time constant set in the context unit activation tends to generate smooth state development with having significant hysteresis. Furthermore, the structures

24

of the hysteresis in the context units are not just arbitrary determined but do through the learning process of the BPTT in which long-time correlations observed in the sensory sequences are extracted as the results of back-propagating the error signals through long time periods. All the characteristics of contextual and gracefully changing behaviors are due to this hysteresis self-organized in the CTRNN.

The idea of switching action primitives by pitchfork bifurcation is analogous to a robot navigation scheme in obstacle environment proposed by Tani and Fukumura (1994). It was shown that robot navigation problems can be simplified by considering branching to left or right only at pitchfork bifurcation points in virtually generated potential fields (Khatib, 1986) using proximity sensors to obstacles. This idea is also analogous to the observations of human rising motions investigated by Yamamoto and Kuniyoshi (2002). Their study showed that diverse trajectories of human rising motions can be qualitatively represented by topological graphs in which each arc denotes a transitional motor act and each node does for a unstable branching point among possible behaviors. In these two studies, the branching structures are mostly generated based on dynamics of physical bodies interacting with environments. Contrary to these studies, the branching structures in the current study are generated based only on the neuro-dynamic structure self-organized through learning. It is more natural if both effects of adaptive neuro-dynamics and physical body dynamics contribute to generations of the branching structures in skilled behaviors of robots. Such investigation is left for future studies.

The results in the current paper can be directly related to the recent finding of the goal encoding IPL neurons by Fogassi et al. (2005). In their experiments, monkeys were trained for two different goal-directed actions. One action is to grasp a piece of food object to eat

and another one is to grasp a solid object (the same size with the food object) to place it into a cylinder. An interesting finding was that activation patterns of many IPL neurons during grasping are different depending on the subsequent goals, namely to eat or to place even though the body kinematics of the monkeys during grasping are the same. Their additional experiments confirmed that the activation preferences during grasping are originated not from the visual stimulus difference between food and solid objects, but from the goal differences. The finding of these goal-specific neurons in IPL may correspond to the current observations for the roles of the context neurons in the CTRNN. We showed that the activation patterns of some context neurons during approaching to the object are specific to the subsequent goals even though the visual and the proprioceptive states profiles during this period are the same. It turned out that these context neurons encode the goals to achieve by utilizing the initial sensitivity characteristics and that their dynamics generates fluent flow of goal-directed skilled behaviors. Although there have been no direct evidences of showing that IPL functions as the sensory forward model and that PMv sets its initial state with goal informations, it is expected that future collaborative studies between empirical neuroscience and synthetics robotics modeling studies would reveal the mechanisms.

Future studies should consider the scaling of the current scheme. We observed that it is difficult to organize action primitives more than 5 and their sequencing dynamics together in the current simple network model. One possible approach is to introduce hierarchy in network architectures. Such trials have been conducted by some of the authors' studies in which gates (Tani & Nolfi, 1999) or so-called the parametric biases (PB) (Tani, 2003) are utilized as interfaces between adjacent levels in the hierarchically organized networks. These additional means such as the gates and the PBs can segment sensory-motor flow into a set

of primitives by utilizing the level structures in the network systems. However, it might be worthwhile to consider the way to scale the current scheme directly without having additional designs such as the gating or the parametric biases. Our preliminary studies have shown that introductions of sparse activations in neuron populations can increase memory capacities of action primitives stored in networks because of reductions of overlaps between activation patterns for different primitives. In the current model, the time constant for context units are set slower than that for sensory units by which the context units dynamics is considered as the higher level and the sensory units as the lower level. These two levels are fully connected by the synaptic weights. It might be worthwhile trying to reduce the connectivities between them because this might reduce unnecessary interferences between the two levels as have been shown in (Paine & Tani, 2005). Our future studies will examine possibilities of the scaling of the current scheme by looking at these points.

7   Summary

The current paper proposed a dynamic neural network scheme which models possible roles of parietal and premotor interactions in acquisition and generation of skilled actions of humans. Our experiments in a object manipulation task showed that the robot can learn to generate multiple goal-directed actions both flexibly and robustly with utilizing the initial sensitivity characteristics of the CTRNN dynamics. Each goal-directed action can be learned with generalization for the object position variances. Furthermore, we observed that the actions can be generated as adaptive to sudden changes of the object position in real time. It was also shown that the robot can recognize the learned goal-directed actions by means of inversely computing the initial context state values. Our analysis showed that the self-organization of

adequate dynamic structures by setting the time constant parameter in appropriate ranges can allow arbitration between two crucial dynamics, namely the top-down goal-directed dynamics and the bottom-up sensory-based dynamics adaptive to the reality. These dynamic structures might be essential for generation of goal-directed skilled actions of humans which should be characterized by generalization, adaptability, robustness and context-dependency.

Acknowledgment

References

Arbib, M. (1981). Perceptual structures and distributed motor control. In *Handbook of Physiology: The Nervous System, II. Motor Control* (pp. 1448–1480). Cambridge, MA: MIT Press.

Beer, R. (1995). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence, 72*(1), 173–215.

Bianco, R., & Nolfi, S. (2004). Evolving the neural controller for a robotic arm able to grasp objects on the basis of tactile sensors. *Adaptive Behavior, 12*(1), 37–45.

Demiris, J., & Hayes, G. (2002). Imitation as a dual-route process featuring predictive and learning components: a biologically plausible computational model. 327–361.

Doya, K., & Yoshizawa, S. (1989). Memorizing oscillatory patterns in the analog neuron network. In *Proc. of 1989 int. joint conf. on neural networks, washington, d.c.* (pp. I:27–32).

Elman, J. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Fogassi, L., Ferrari, P., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science, 308*, 662–667.

Heilman, K. (1973). Ideational apraxia - a re-definition. *Brain, 96*, 861–864.

Ito, M., & Tani, J. (2004). On-line imitative interaction with a humanoid robot using a dynamic neural network model of a mirror system. *Adaptive Behavior, 12*(2), 93–114.

Jordan, M. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. In *Proc. of eighth annual conference of cognitive science society* (pp. 531–546). Hillsdale, NJ: Erlbaum.

Kawato, M., Furukawa, K., & Suzuki, R. (1987). A hierarchical neural network model for the control and learning of voluntary movement. *Biological Cybernetics*, *57*, 169–185.

Khatib, O. (1986). Real-time obstacle avoidance for manipulators and mobile robots. *The Int. J. of Robotics Research*, *5*(1), 90–98.

Liepmann, H. (1920). Apraxie. *Erg ges Med*, *1*, 516–543.

Luria, A. (1973). *The working brain.* Penguin Books Ltd.

Nishimoto, R., & Tani, J. (2004). Learning to Generate Combinatorial Action Sequences Utilizing the Initial Sensitivity of Deterministic Dynamical Systems. *Neural Networks*, *17*, 925–933.

Oztop, E., & Arbib, M. A. (2002). Schema Design and Implementation of the Grasp-Related Mirror Neuron System. *Biological Cybernetics*, *87*, 116–140.

Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, *22*, 129–151.

Paine, R., & Tani, J. (2005). How hierarchical control self-organizes in artificial adaptive systems. *Adaptive Behavior*, *13*(3), 211–225.

Rizzolatti, G., Fadiga, L., Galless, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131–141.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing* (pp. 318–362). Cambridge, MA: MIT Press.

Tani, J. (1996). Model-Based Learning for Mobile Robot Navigation from the Dynamical Systems Perspective. *IEEE Trans. on SMC (B)*, *26*(3), 421–436.

Tani, J. (2003). Learning to generate articulated behavior through the bottom-up and the top-down interaction process. *Neural Networks*, *16*, 11–23.

Tani, J., & Fukumura, N. (1994). Learning Goal-directed Sensory-based Navigation of a Mobile Robot. *Neural Networks*, *7*(3).

Tani, J., & Ito, M. (2003). Self-organization of behavioral primitives as multiple attractor dynamics: a robot experiment. *IEEE Trans. on Sys. Man and Cybern. Part A*, *33*(4), 481–488.

Tani, J., Ito, M., & Sugita, Y. (2004). Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. *Neural Networks*, *17*, 1273–1289.

Tani, J., & Nolfi, S. (1999). Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems. *Neural Networks*, *12*, 1131–1141.

Wiggins, S. (1990). *Introduction to Applied Nonlinear Dynamical Systems and Chaos.* New York: Springer-Verlag.

Williams, R. J., & Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, *1*, 270–280.

Wolpert, D., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*, 1317–1329.

Yamamoto, T., & Kuniyoshi, Y. (2002). Stability and controllability in a rising motion: a global dynamics approach. In *Proceedings of ieee international conference on intelligent robots and system 2002* (pp. 2467–2472).
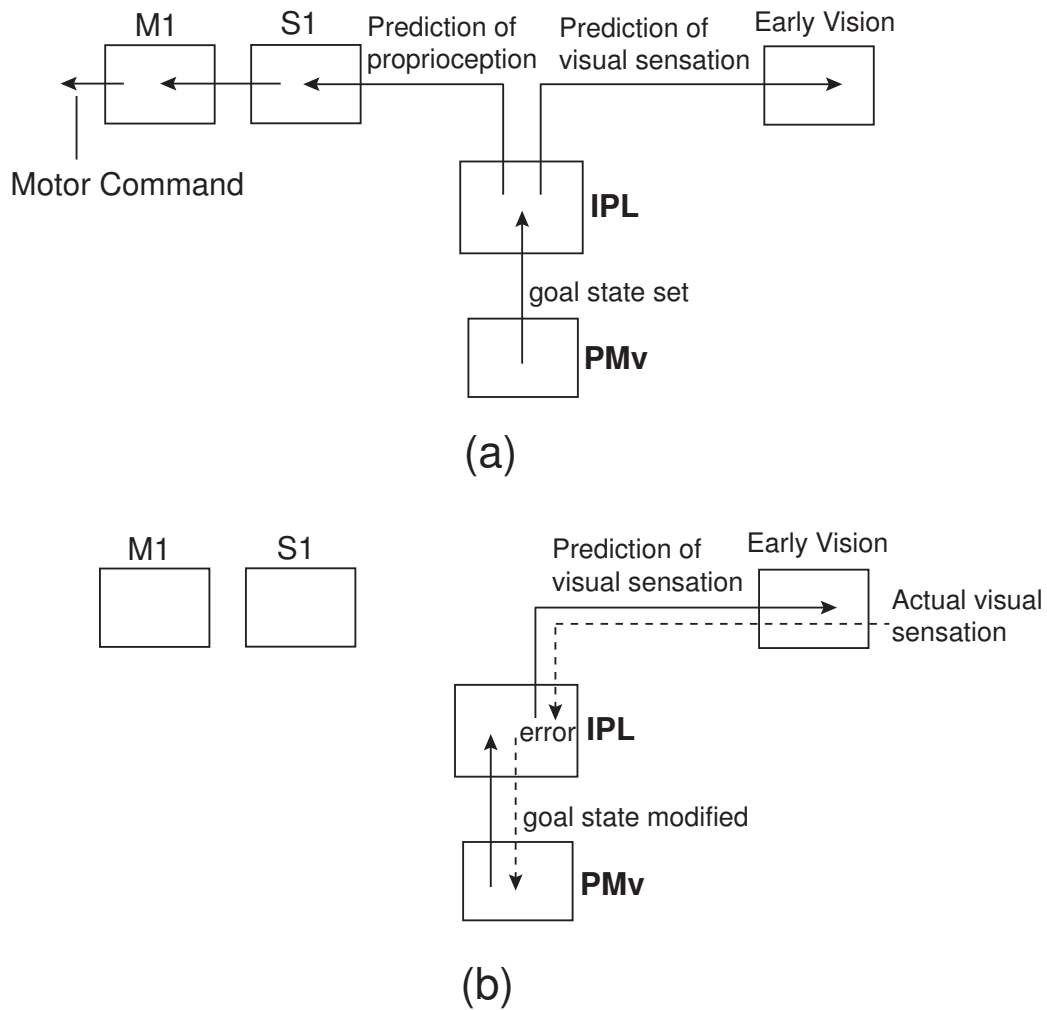
Figure 1: (a)The action generation pathway. PMv sends the goal information to the sensory forward model in IPL and the sensory forward model generates the predicted sensory sequence. The predicted proprioceptive state is sent to S1 and further to M1 where the corresponding motor command is generated. (b)The action recognition pathway. The error between the predicted visual image and the actual one is back-propagated to PMv and the goal state is modified.
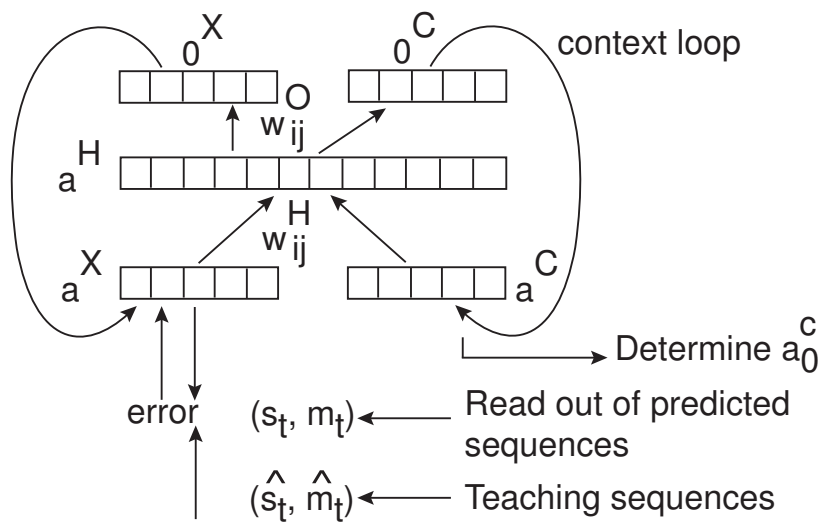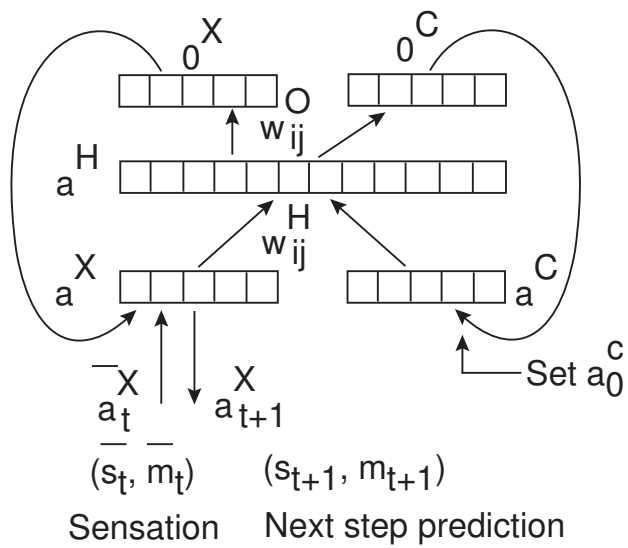
Figure 2: The CTRNN model employed. (a) training mode and (b) behavior generation by open-loop mode.
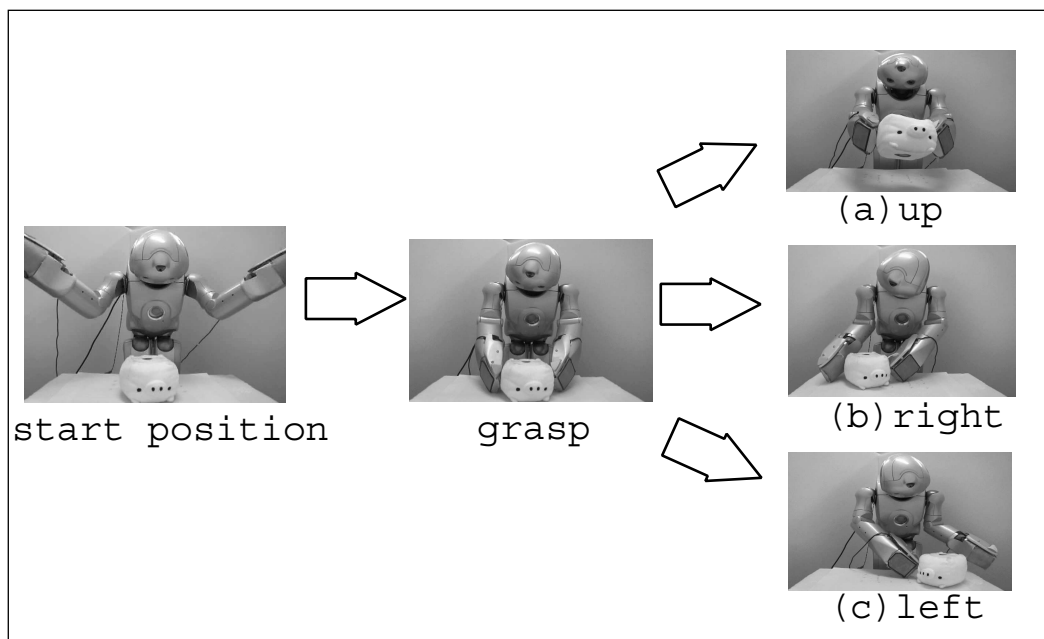
Figure 3: Three type of the goal-directed behaviors aimed in the experiment using a SONY produced humanoid robot.
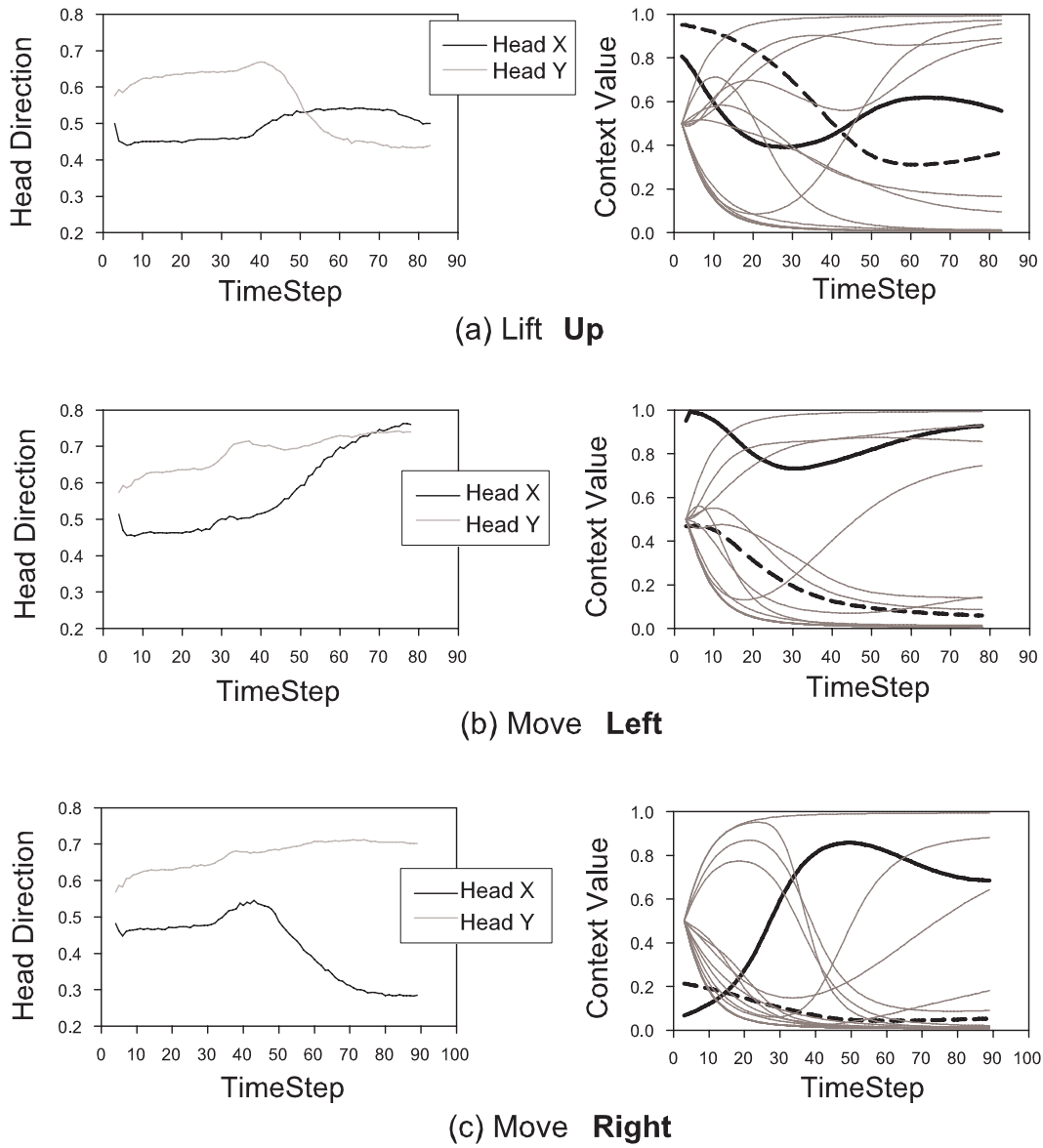
Figure 4: The generation of (a) "lift up", (b) "move left" and (c) "move right" actions with the object allocated in the center. In left-hand side, head direction represents relative position of the object to be manipulated. In right-hand side, a bold curve and a dotted one represent two context unit activation values of which initial states are adapted.

37
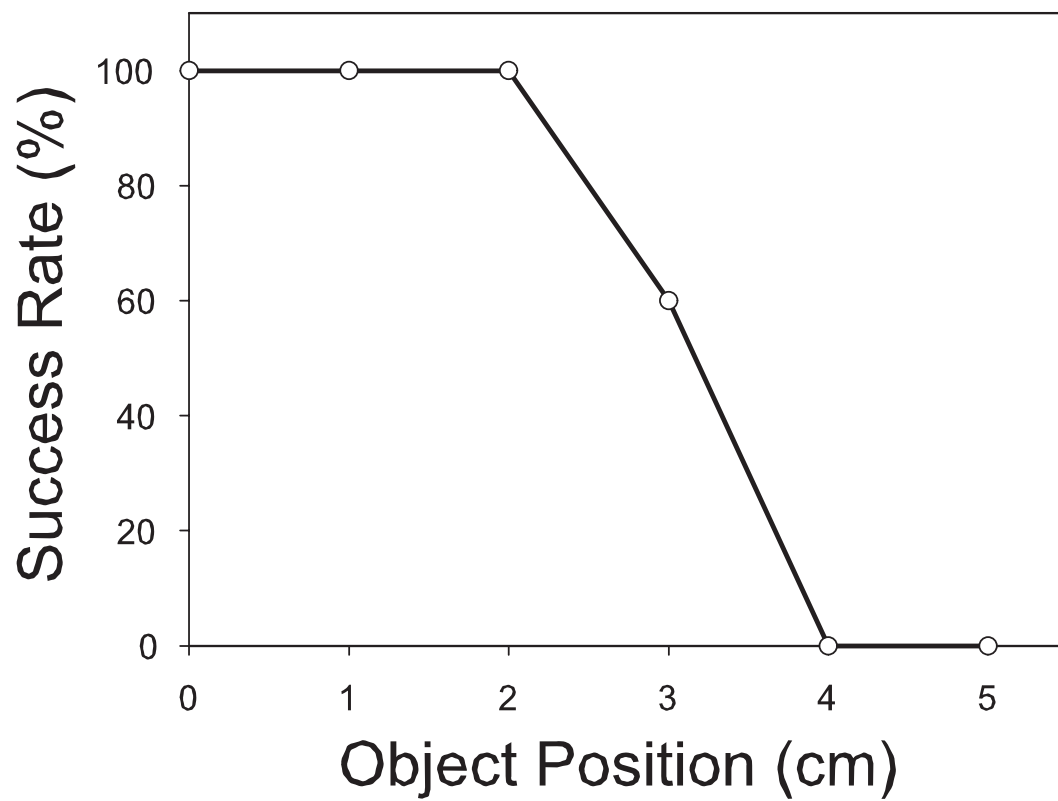
Figure 5: The success rate of the lift up action with varying the object position to further left of the trained range.
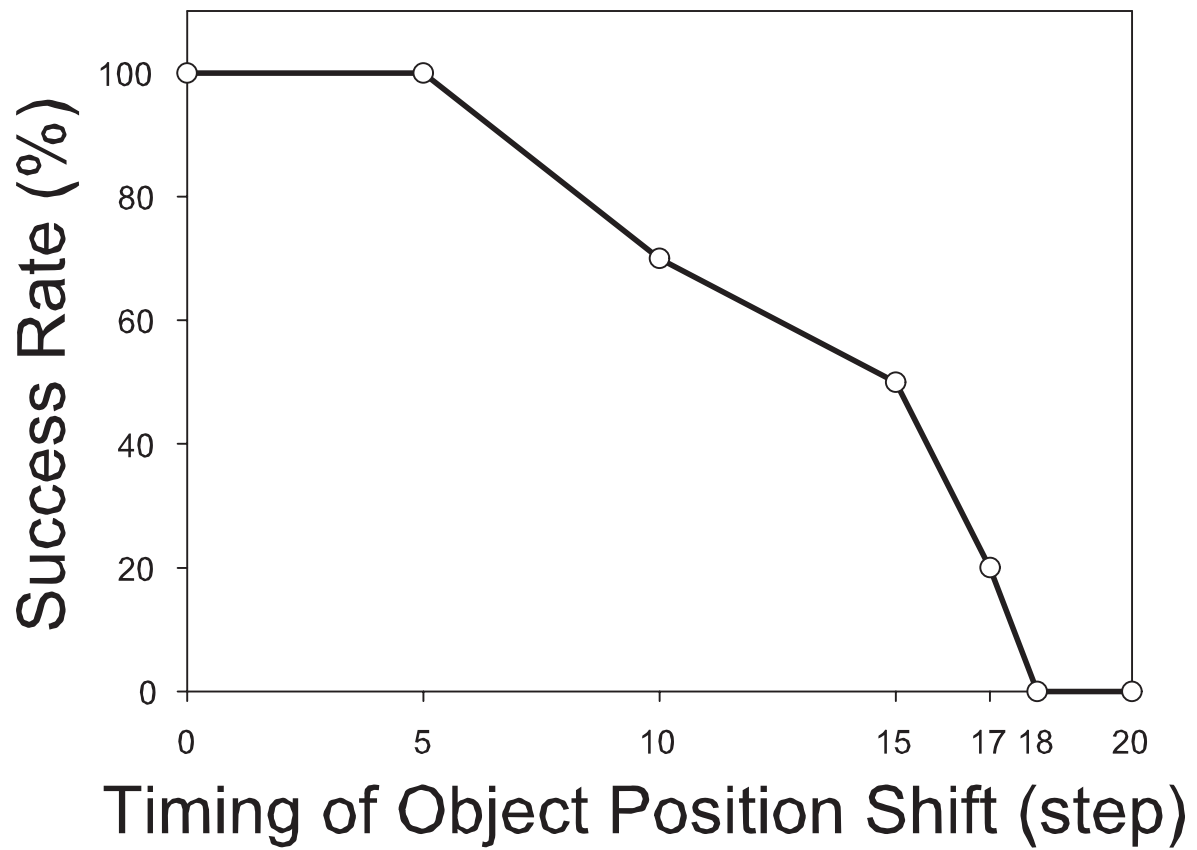
Figure 6: The success rate of the lift up action with varying the timing of object position shifts.
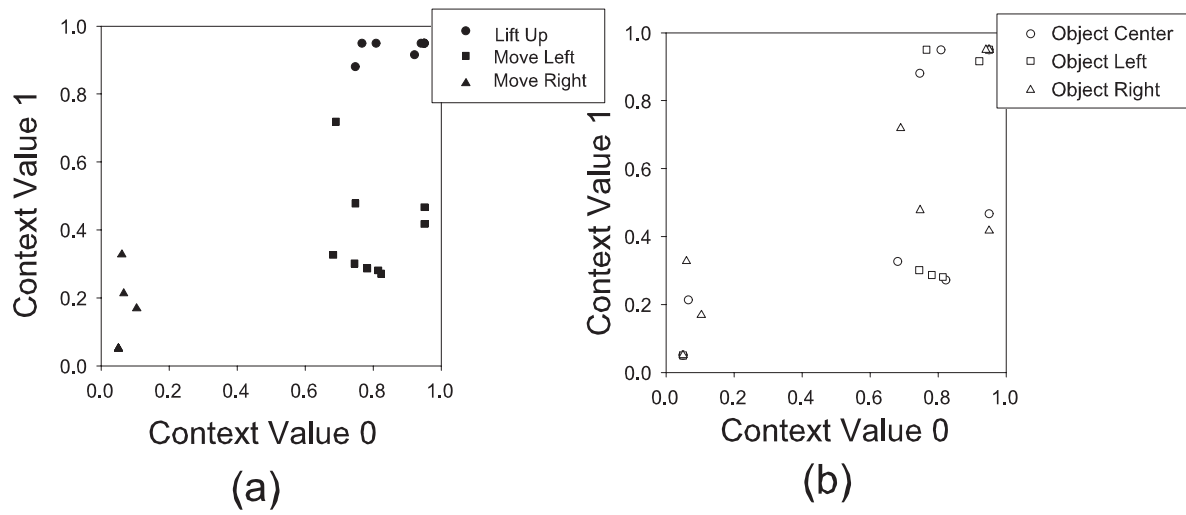
Figure 7: The initial context state map with category labels denoted for (a) different goals and (b) for different object positions.
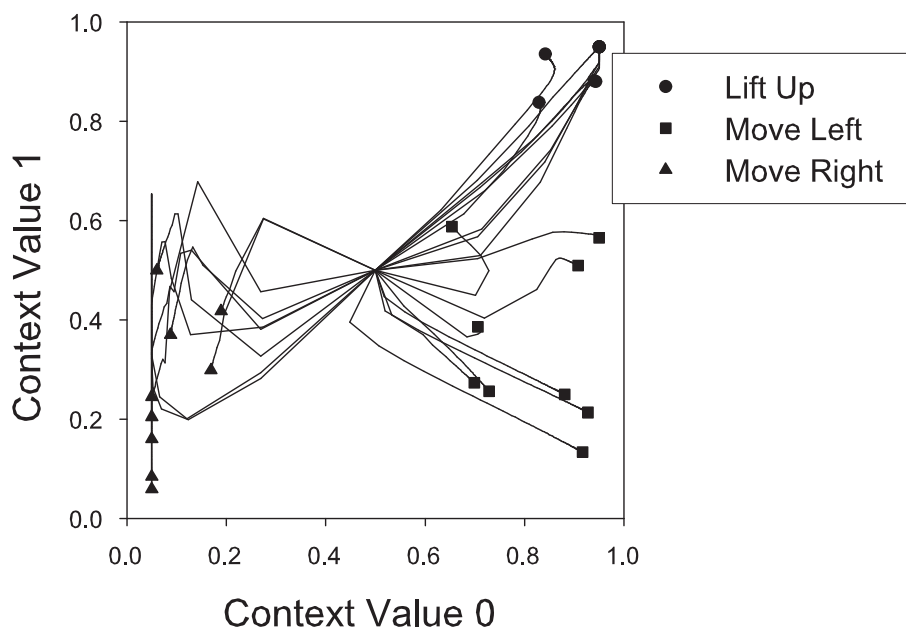
Figure 8: The trajectories of searching the optimal value of the initial context state in recognizing the observed different goal-directed behaviors. The black category labels denote the converged values.
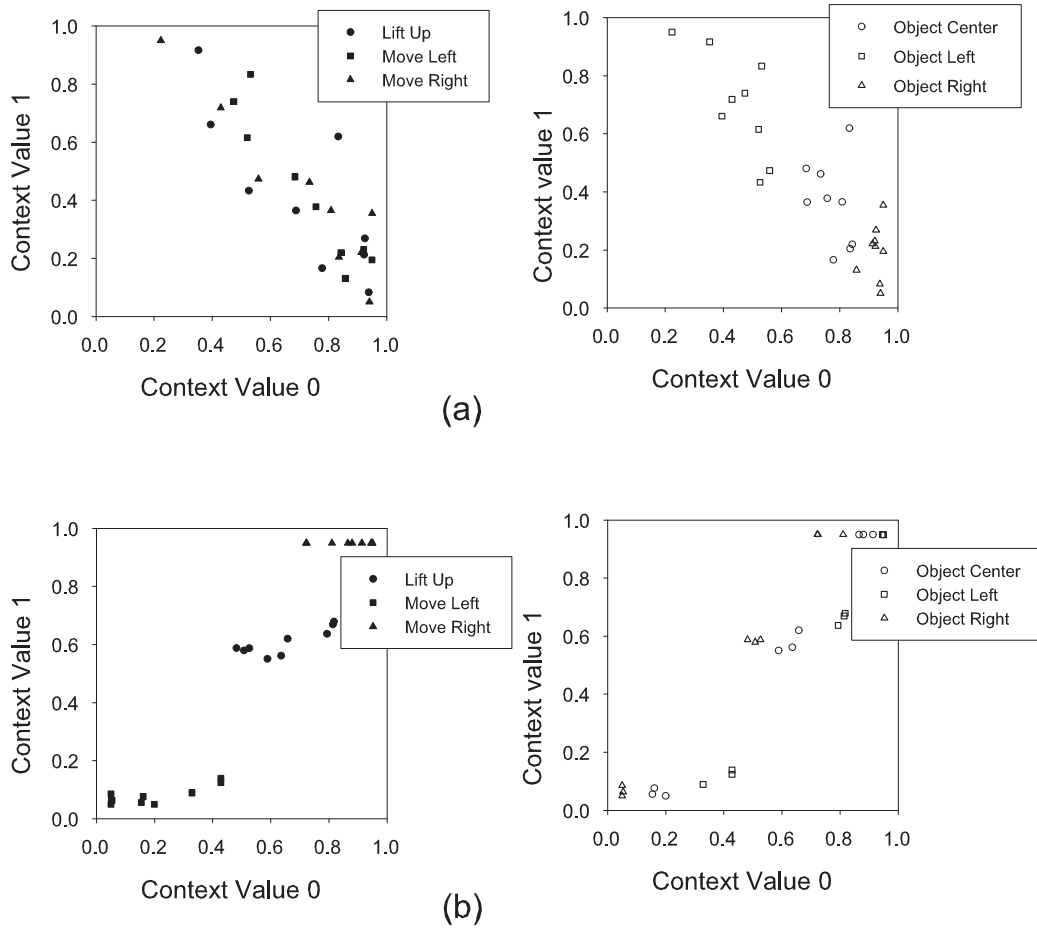
Figure 9: The initial context state map self-organized with smaller time constant parameters (a) and larger time constant parameters (b) with category labels denoted for different goals in left-hand side and different object positions in the right-hand side.
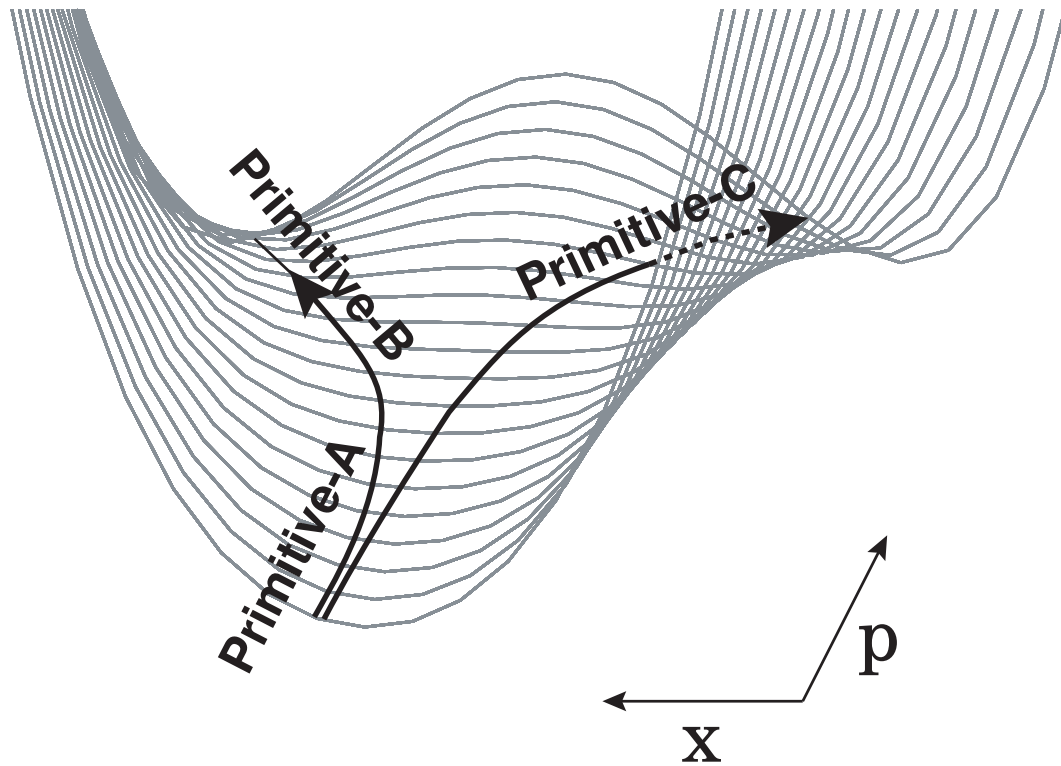
Figure 10: The branching by Pitchfork bifurcation. Initially there is only a stable fixed point in the direction of x, representing action primitive A. As P develops in time, the fixed point bifurcates into two stable fixed points, representing action primitive B and action primitive C.