Developmental Dynamics of RNNPB: New Insight about Infant Action Development

Jun-Cheol Park[1], Dae-Shik Kim[1], and Yukie Nagai[2]

[1] Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea
{pakjce,daeshik}@kaist.ac.kr
[2] Department of Adaptive Machine Systems, Osaka University, Osaka 565-0871, Japan
yukie@ams.eng.osaka-u.ac.jp

Abstract. Developmental studies have suggested that infants' action is goal-directed. When imitating an action, younger infants tend to reproduce the goal while ignoring the means (i.e., the movement to achieve the goal) whereas older infants can imitate both. We suggest that the developmental dynamics of a Recurrent Neural Network with Parametric Bias (RNNPB) may explain the mechanism of infant development. Our RNNPB model was trained to reproduce six types of actions (2 different goals x 3 different means), during which parametric biases were self-organized to represent the difference with respect to both the goal and means. Our analysis of the self-organizing process of the parametric biases revealed an infant-like developmental change in action learning: the RNNPB first adapted to the goal and then to the means. The different saliency of these two features caused this phased development. We discuss the analogy of our result to infant action development.

Keywords: Cognitive developmental robotics,Infant action development, Recurrent neural network, RNNPB.

1 Introduction

It is known that infants can understand and imitate adults' goal-directed actions as reported in previous empirical studies [1,2]. The study of Carpenter et al. [3] compared an ability of imitation of a goal-directed behavior between 12-month-old and 18-month-old infants when an adult demonstrated actions with two different goals and two different motion styles. Their results demonstrated that, younger infants achieved the goals of the actions while ignoring the motion styles (i.e., the means). Older infants, however, reproduced the both without ignoring them.

Why is there a difference in the ability between younger and older infants in terms of two aspects of the goal-directed action? A perspective of cognitive developmental robotics has suggested computational approaches to understanding the internal mechanisms of cognitive developmental process of humans [4,5].

A.P. del Pobil et al. (Eds.): SAB 2014, LNAI 8575, pp. 144–153, 2014.

Tani and Ito [6] suggested a neural network model called a Recurrent Neural Network with Parametric Bias (RNNPB). The key feature of an RNNPB is that it can encode multiple dynamic patterns into a static activity of the parametric bias (PB) units and the representation of the PB units is self-organized during learning process. It is also known that the RNNPB has biological plausibility such as mirror neuron property [6, 7]. Hence, this architecture has been used in robotic experiments for goal-directed action imitation [6, 8]. The study by Ito et al. [9] showed that a self-organized representation of the PB units exhibits a generalization capability in case of simple action learning.

Our study investigates how PB units are gradually self-organized during learning of an RNNPB in order to represent the two aspects of goal-directed actions (i.e., the goal and means). It can be expected that the PB units would be organized to be able to distinguish all trained actions, as previous studies have shown. The dynamical changes in learning process, however, cannot be predicted because there have not yet been any studies of that. If there are meaningful relation between the dynamic changes of the PB units and the two aspects of goal-directed actions, it would provide new insights into the mechanism of infant development of goal-directed behavior.

2 Goal-Directed Behavior

A virtual robot arm, which consists of two joints, is defined in a simulated environment as illustrated in Fig.1(a). In this environment, each joint ($\boldsymbol{\theta} = [\theta_1, \theta_2]^T$) moves from 0 to 180 degrees in a two dimensional space. Inspired by the experiment of Carpenter et al. [3], goal-directed actions are designed to reveal two aspects of reaching behavior: the goal and the means. The goal of an action is moving the arm from the initial position to one of two goal positions (A, B). The means of an action is matching the trajectory of the movement.

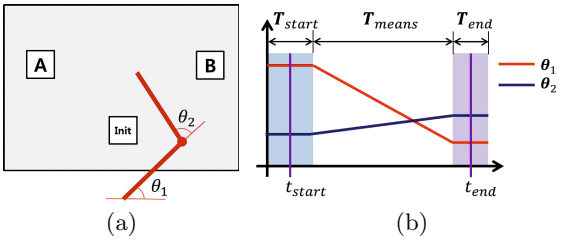

Fig. 1. (a) An overview of task and simulation environment. (b) The sequence of joint angles.

As shown in Fig. 1(b), the motor behavior consists of three parts. In the first part, the arm waits at the initial state for T_{start} time steps, and then moves to the goal state for T_{means}. Finally, it stays at the goal state for T_{end}. Hence, a

Table 1. Lists of reference behaviors

	Goal	Period	Amplitude		Goal	Period	Amplitude
$\mathbf{A_0}$	A	-	-	$\mathbf{B_0}$	B	-	-
$\mathbf{A_1}$	A	T_{means}	α	$\mathbf{B_1}$	B	T_{means}	α
$\mathbf{A_2}$	A	$0.5T_{means}$	α	$\mathbf{B_2}$	B	$0.5T_{means}$	α

reaching behavior appears in the form of time sequence of joint angles $\mathbf{\Theta}$ whose length is $L = T_{start} + T_{means} + T_{end}$.

For each goal position (A, B), there are three different types of movements to achieve the goal (see Fig.2). Thus, there are totally six reference motor behaviors (2 simple movements + 4 hopping movements) as explained in Table 1. A_0 and B_0 are simple movements, which have straight trajectories of the joint angles from the initial posture to the goal posture (see Figs.2(a) and (d)). A_1, A_2, B_1 and B_2 are, in contrast, hopping movements, which add a sinusoidal perturbation with a different period (see Figs. 2(b), (c), (e) and(f)). In our experiment, the agent is supposed to experience a desired motor behavior $\mathbf{\Theta_{ref}} \in \left\{\mathbf{\Theta}_{\{A_0,A_1,A_2\}}, \mathbf{\Theta}_{\{B_0,B_1,B_2\}}\right\}$ through, for example, kinesthetic teaching.

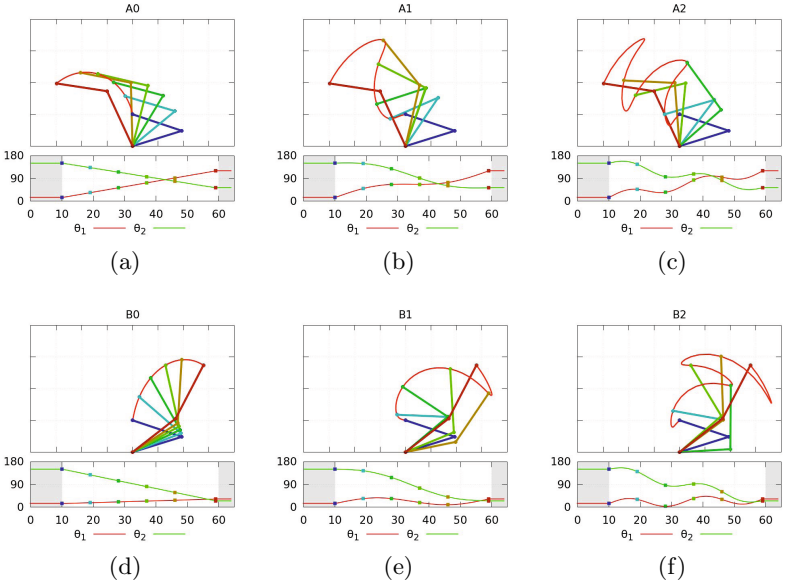

Fig. 2. Three different movements for two goal positions. (a) to (c) are reaching for the goal A, and (d) to (f) are for the goal B.

3 An RNNPB Model for Learning Agent

The main feature of an RNNPB [6] is that it can encode multiple dynamics of input/output relationships using static activation of PB units, where PB units are self-organized through learning process. The agent can be modeled to build its own internal memory for goal-directed actions based on its own experiences.

3.1 Architecture of the Model

As modified version of Jordan-type recurrent neural networks [10], an RNNPB basically consists of input/output units in a three-layered structure and context units with closed feedback. In addition, it has PB units in the input layer, which enable the network to learn multiple actions. (see Fig. 3(a)). The parameters of the network are $\Psi = \{\mathbf{W}_{21}, \mathbf{W}_{32}, \mathbf{b}_2, \mathbf{b}_3\}$, where $\mathbf{W}_{21}$ and $\mathbf{b}_2$ are the connecting weight and the bias between the first and hidden layers, and $\mathbf{W}_{32}$ and $\mathbf{b}_3$ are the same between the hidden and third layers.

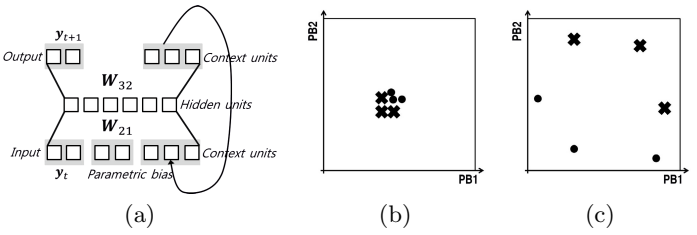

Fig. 3. (a) An RNNPB model consists of three layers: input layer, hidden layer, and output layer. (b) and (c) indicate the PB space before learning and after learning, respectively. The six goal-directed actions $\mathbf{Y}_{ref}$ are encoded as different PB values $\mathbf{x}_{recog}$, which are illustrated as markers in the space.

In our experiment, the activity of input/output units represent normalized joint angles of the virtual robotic arm $\boldsymbol{\theta}$. Hence the number of input/output units is 2, and the activity of input/output units is denoted by $\mathbf{y} = [y_1, y_2]^T$. PB units have two elements because it is easy to visualize and analyze them in a two-dimensional space. The number of the context and hidden units is empirically set to be able to represent all the reference behaviors.

3.2 Learning Procedure

The main rule of learning is to update the network parameters Ψ to enable the agent to generate desired motor behaviors. The agent initially has randomly initialized network parameters Ψ^0. When the agent experiences a desired motor behavior $\mathbf{Y_{ref}} \in \left\{\mathbf{Y}_{A_{\{0,1,2\}}}, \mathbf{Y}_{B_{\{0,1,2\}}}\right\}$ as a form of a normalized time sequence $\mathbf{Y} = [\hat{\mathbf{y}}_1, \cdots, \hat{\mathbf{y}}_L]$, the network parameters Ψ^n at the n-th iteration are updated

to minimize errors $\mathbf{E}_t^{out}$ between the desired values $\hat{\mathbf{y}}_t$ and the outputs predicted by the network $\mathbf{y}_t$ for all time steps ($1 \leq t \leq L$). Back Propagation Through Time (BPTT) [11] is used for updating Ψ.

$$\mathbf{E}_t^{out} = \hat{\mathbf{y}}_t - \mathbf{y}_t \tag{1}$$

During the BPTT process, the back-propagated delta values for the hidden units δ_{hid}, the input units δ_{in}, the output units δ_{out}, the context units δ_{cxt}, and the PB units δ_{PB} are calculated from the error of the outputs $\mathbf{E}_t^{out}$ from 1 to T time steps. Especially values of PB units ($\mathbf{x} = [x_1, x_2]^T$) are updated by Eq. (2).

$$\Delta\rho_{PB}^{(i)} = k_{bp} \sum_{t=1}^{T} \delta_{PB,t}^{(i)}$$

$$x^{(i)} = sigmoid(\rho_{PB}^{(i)}) \tag{2}$$

After the value of the PB units $\mathbf{x}_i^{(n)}$ is determined for behavior i at the n-th iteration, it is used for calculating the forward dynamics and updating ψ^{n+1} in the $(n+1)$-th iteration.

3.3 Recognition and Generation Procedure

When the normalized reference behaviors $\mathbf{Y}_{ref}$ are given in the recognition phase, the RNNPB firstly finds corresponding PB values $\mathbf{x}_{recog}$ based on the BPTT algorithm. Unlike the learning procedure, only PB values are updated in the recognition phase.

In the generation phase, the RNNPB generates new motor behaviors $\mathbf{Y}_{gen}$ based on the PB values $\mathbf{x}_{recog}$ recognized in the previous phase. The activity of the output units at t time step is calculated using the network parameters Ψ. As the output is directly used as the values of the input units at $t+1$ time step, a sequence of motor behavior is generated step by step.

3.4 Role of PB Values

A set of static PB values represents a corresponding action as a characteristic of the RNNPB model. The PB values, which range from 0.0 to 1.0, are self-organized through learning so as to represent all the experienced actions. Thanks to linearity (not globally but locally) in the PB space, the RNNPB can also represent novel behaviors by combining experienced actions. Figs. 3(b) and (c) illustrate how the PB space is self-organized through learning. Before learning (see Fig. 3(b)), the six goal-directed actions are not separated from each other. The RNNPB at this stage would thus produces only one type of motion. The PB values for the six actions gradually change during learning. After learning (see Fig. 3(c)), the six actions will be discriminated as six different PB values, which can generate the desired actions.

4 Experimental Results

We trained an RNNPB model to reproduce the six goal-directed actions described in Section 2.

4.1 Learning Curve

The ability of an agent with the RNNPB model was assessed in terms of two viewpoints. The first point of view is whether the agent successfully reaches the desired goal posture from the initial posture. An error E_{goal} was calculated by taking the average of two error values at the initial posture E_{start} and the end posture E_{end}. The two error values (E_{start} and E_{end}) were obtained by calculating Euclidian distance between the reference motor behaviors $\mathbf{y}_t^{ref} \in \mathbf{Y}_{ref}$ and generated motor behaviors $\mathbf{y}_t^{gen} \in \mathbf{Y}_{gen}$ at t_{start} and t_{end}, respectively (see Fig. 1(b)).

$$E_{start} = \left\| \mathbf{y}_{t_{start}}^{ref} - \mathbf{y}_{t_{start}}^{gen} \right\|, \quad E_{end} = \left\| \mathbf{y}_{t_{end}}^{ref} - \mathbf{y}_{t_{end}}^{gen} \right\|$$

$$E_{goal} = \frac{E_{start} + E_{end}}{2} \tag{3}$$

The second point of view is measuring how accurately the agent traces the style of movements. An error of the means E_{means} was defined as the averaged error over T_{means} time steps, where Euclidian distance between the reference motor behaviors $\mathbf{y}_t^{ref}$ and generated ones $\mathbf{y}_t^{gen}$ was applied.

$$E_{means} = \frac{1}{T_{means}} \sum_{t \in \mathcal{T}_{means}} \left\| \mathbf{y}_t^{ref} - \mathbf{y}_t^{gen} \right\| \tag{4}$$

Fig. 4 shows the average value of E_{goal} and E_{means} for 100 RNNPBs with different initial parameters Ψ^0. As the RNNPB models have been trained, the error values also have decreased. The average error for the goal became smaller than the average error of the means when the networks had been trained enough.

4.2 Dynamics of PB Space and Generated Output

To investigate the developmental dynamics of the PB space and its relation to the action generation, E_{means} was examined for all possible PB values. Additionally a recognized PB value $\mathbf{x}_{ref}$ for each reference action $\mathbf{Y}_{ref}$ and generated output $\mathbf{Y}_{gen}$ were calculated. Three iteration points (0, 10,000 and 200,000) were picked up based on the error curves to show the dynamical self-organization of the PB space. Fig. 5 shows the result for an RNNPB among the 100 trained RNNPBs. On the left-side of this figure (see Figs. 5 (a), (c) and (e)), the direction and the

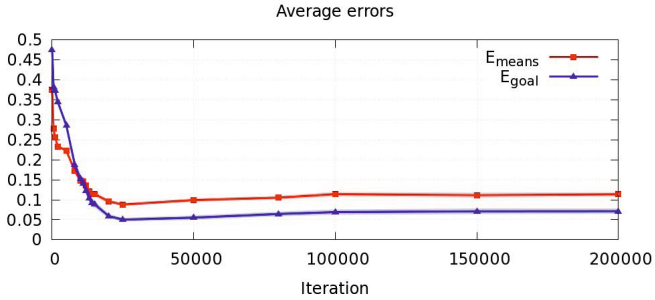

Fig. 4. The transition of errors in terms of the goal E_{goal} and the means E_{means}. The two curves plot the average of 100 networks with different initial parameters.

color of triangular markers indicate which types of actions ($\mathbf{S_x}$) have a minimum error value with the corresponding PB values.

$$\mathbf{S_x} = \underset{\mathbf{S}\in\{A_0,\cdots,B_2\}}{\operatorname{argmin}} \left\{ E_{means} \middle| \mathbf{Y}_{ref=\mathbf{Y_S}} \right\} \tag{5}$$

On the right-side of the figure (see Figs. 5 (b), (d) and (f)), the reference behaviors $\mathbf{Y}_{ref}$ (thick lines) and the generated outputs $\mathbf{Y}_{gen}$ (thin lines) are plotted for the six different actions. Recognized PB values for each reference behavior $\mathbf{x}_{recog}$ are painted as a circle with a triangular marker in (a), (c), and (e).

The result shows that the agent has gradually improved the ability to reproduce the reference actions as it increases experiences. Meanwhile, the PB space is gradually self-organized to represent the actions. When the agent has no experience of the reference behaviors (0 iteration), it cannot produce the desired actions due to the undifferentiated PB values (see Figs. 5 (a) and(b)). When the agent has been trained for 10,000 iterations, it generates simpler behaviors (i.e. A_0 and B_0) well, while producing larger errors for the hopping actions (i.e., A_1, A_2, B_1, and B_2) (see Figs. 5 (c) and (d)). The PB space separates only between A and B but not within A and B, indicating that the goal of the actions has been acquired but the means has not yet. When the agent is fully trained, it finally generates the six actions in terms of both the goal and the means. The well-organized PB values enable the agent to discriminate the actions (see Figs. 5 (e) and (f)). Taken all together, phased learning (i.e., first learning the goal of the actions and then the means) has been achieved through the development.

5 Discussion

Empirical studies of infant action development have shown that only older infants can imitate both of two aspects of the adults goal-directed actions. Our result of the self-organizing dynamics of the PB space also showed similar characteristic of the infant development. In terms of the trajectory of the reference behavior

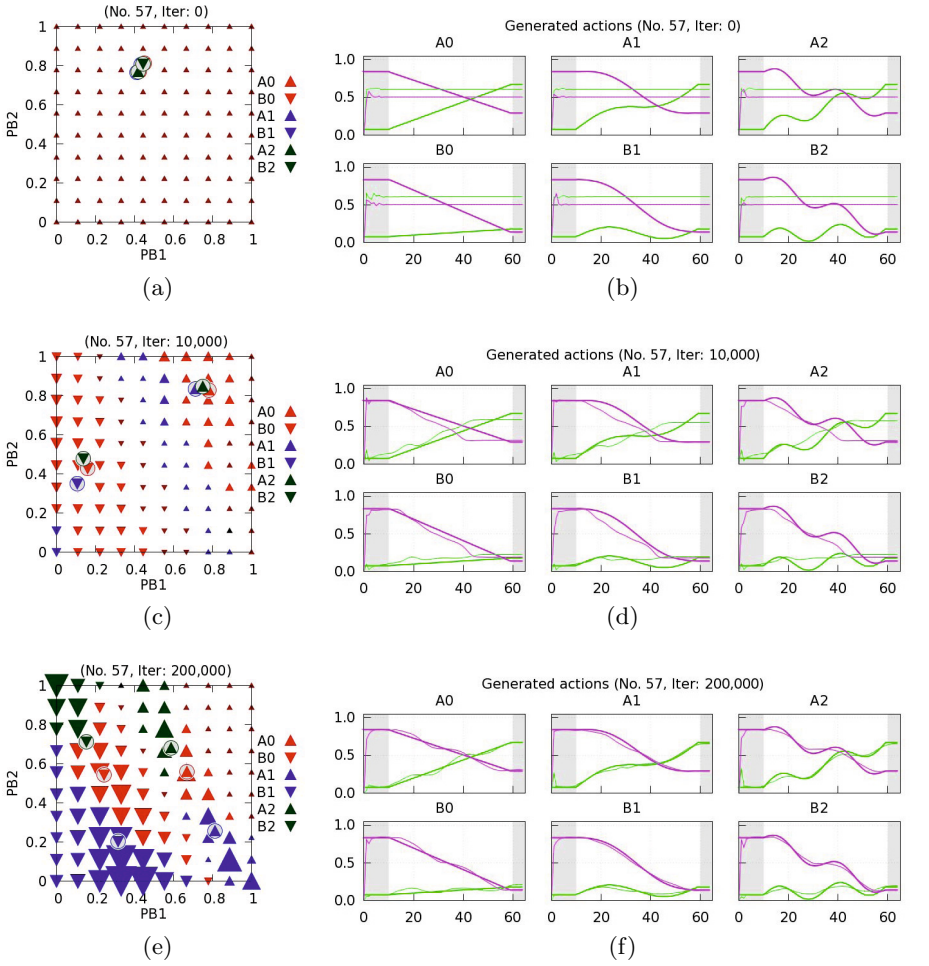

Fig. 5. Dynamics of PB space and results of action generation. (a) (c) and (e) illustrate which reference actions (from A_0 to B_2) have a minimum error in the PB space. The direction and the colors of triangular markers indicate the goal and the style of motion, respectively. The size of the markers indicates the amount of error E_{means}: The larger a marker is, the smaller the error is. Recognized PB values $\mathbf{x}_{recog}$ are illustrated as a circle with triangular markers inside. (b) (d) and (f) represent $\mathbf{Y_{gen}}$ (thin lines) for all reference actions $\mathbf{Y_{ref}}$ (thick lines) in the time domain. The red and green lines are the first and second joint angles, respectively.

$\mathbf{Y}_{ref}$ and the generated action $\mathbf{Y}_{gen}$ (see Fig. 5), the immature agent that was trained 10,000 iterations reproduced the goal position well and only the simple trajectories without sinusoid movement. The mature agent, on the other hand, reproduced both of the goal postures and the shapes of the trajectories. As a result, it was found that the RNNPB model first adapted to the goal and then the means among the two aspects of motions through learning process.

As reported in the studies by Carpenter et al. [3] and Bekkering et. al [12], infants imitate adults action differentially based on the salience of the actions. For instance, Carpenter et al. [3] compared two different types of movements: (a) sliding an object and (b) bouncing the object several times. Both 12- and 18-month-old infants tended to ignore the means of motions when a toy house was given as the goal position (House condition). In contrast, when the goal position was not indicated (No house condition), infants imitated well the means of the motions (sliding and hopping) as ignoring the goals.

In case of the RNNPB model, the error function used for the BPTT learning (see Eq. 2) seems to be a key rule for the phased development of the two aspects of the actions. The internal parameter of the network Ψ is updated at one iteration and the amount of the update is calculated based on the error values for the whole action sequence and for all the reference behaviors. Hence, the effects of error values are averaged for all the reference behaviors. That is, it makes the error of the transition part (i.e., the means) diminish while relatively enhancing the error in the initial and the goal states of the actions. The errors of the means, however, become a salient feature after the errors of the goal decreased enough. Another characteristic of the RNNPB model is that more than one behavior can be encoded into the PB values by the one network parameter.

An interesting point is that the change of the salient features of actions in the RNNPB model is due to its internal maturing procedure without changing the capacity of the network or giving any external signal. This could be one of possible explanations for the development of infants' ability for goal-directed actions. However, there is still a small gab between our model and infant mechanism in terms of the representation of behaviors. Therefore, we intend to improve our model by including multimodal representation such as visual information and to examine whether it can better simulate infant experiments.

6 Conclusion

In this study, we trained the RNNPB model to reproduce the six goal-directed actions using the virtual robot arm. The goal-directed actions were composed of the two different goal positions and the three different means of motions. While the network was gradually trained, the organization of the PB space and the generated actions were analyzed. As a result, the agent trained for 10,000 iterations could generate the simple actions well as fulfilling the goal but not the hopping actions. The agent trained for 200,000, in contrast, could generate both the simple and the hopping actions accurately. Taken together, our RNNPB model first adapted the goal and then the means during learning process. Finally,

we discussed that these self-organized developmental changes in the RNNPB may explain the mechanism of infant development of goal-directed actions.

Acknowledgment. This study is partially supported by JSPS/MEXT Grants-in- Aid for Scientific Research (Research Project Number: 24000012, 24119003, 25700027), by the Research Center Program of IBS(Institute for Basic Science) in Korea(HQ1201) and by the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planing (NRF-2010-0018837).

References

1. Meltzoff, A.N.: Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. Developmental Psychology 31(5), 838 (1995)
2. Carpenter, M., Akhtar, N., Tomasello, M.: Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. Infant Behavior and Development 21(2), 315–330 (1998)
3. Carpenter, M., Call, J., Tomasello, M.: Twelve-and 18-month-olds copy actions in terms of goals. Developmental Science 8(1), F13–F20 (2005)
4. Asada, M., MacDorman, K.F., Ishiguro, H., Kuniyoshi, Y.: Cognitive developmental robotics as a new paradigm for the design of humanoid robots. Robotics and Autonomous Systems 37(2), 185–193 (2001)
5. Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., Yoshida, C.: Cognitive developmental robotics: A survey. IEEE Transactions on Autonomous Mental Development 1(1), 12–34 (2009)
6. Tani, J., Ito, M., Sugita, Y.: Self-organization of distributedly represented multiple behavior schemata in a mirror system: Reviews of robot experiments using rnnpb. Neural Networks 17, 1273–1289 (2004)
7. Cuijpers, R.H., Stuijt, F., Sprinkhuizen-Kuyper, I.G.: Generalisation of action sequences in rnnpb networks with mirror properties (2009)
8. Yokoya, R., Ogata, T., Tani, J., Komatani, K., Okuno, H.G.: Experience-based imitation using rnnpb. Advanced Robotics 21(12), 1351–1367 (2007)
9. Ito, M., Tani, J.: Generalization in learning multiple temporal patterns using rnnpb. In: Pal, N.R., Kasabov, N., Mudi, R.K., Pal, S., Parui, S.K. (eds.) ICONIP 2004. LNCS, vol. 3316, pp. 592–598. Springer, Heidelberg (2004)
10. Jordan, M.: Attractor dynamics and parallelism in a connectionist sequential network. In: Proceedings of the Eighth Annual Conference of the Cognitive Science Society (1986)
11. Werbos, P.J.: Backpropagation through time: What it does and how to do it. Proceedings of the IEEE 78(10), 1550–1560 (1990)
12. Bekkering, H., Wohlschlager, A., Gattis, M.: Imitation of gestures in children is goal-directed. The Quarterly Journal of Experimental Psychology: Section A 53(1), 153–164 (2000)