2012 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 7-12, 2012. Vilamoura, Algarve, Portugal

Perceptual Development Triggered by its Self-Organization in Cognitive Learning

Yuji Kawai, Yukie Nagai, and Minoru Asada

Abstract— **It has been suggested that perceptual immaturity in early infancy enhances learning for various cognitive functions. This paper demonstrates the role of visual development triggered by self-organization in a learner's visual space in a case of the mirror neuron system (MNS). A robot learns a function of the MNS by associating self-induced motor commands with observed motions while the observed motions are gradually self-organized in the visual space. A temporal convergence of the self-organization triggers visual development, which improves spatiotemporal blur filters for the robot's vision and thus further advances self-organization in the visual space. Experimental results show that the self-triggered development enables the robot to adaptively change the speed of the development (i.e., slower in the early stage and faster in the later stage) and thus to acquire clearer correspondence between self and other (i.e., the MNS).**

I. Introduction

Human infants are born with immature abilities. Developmental psychologists have suggested that such immature functions facilitate the acquisition of higher cognitive functions, which is called maturational constraints [1]–[3]. However, it is difficult to reveal the role of maturational constraints by comparing infants with and without such constraints since every infant has immature abilities.

Recently, computational studies demonstrating advantages of maturational constraints have attracted researchers' attention [4]–[8]. Computational models allow researchers to compare the performance of learning with and without maturational constraints. Elman [4] verified that limited capacity of infants' memory facilitates language learning. Dominguez and Jacobs [5] proposed a model for learning of binocular disparity with visual development. Furthermore, it has been shown that perceptual and motor development triggered by learning progress achieves faster and more generalized learning than pre-scheduled development [6]–[8]. Nagai et al. [6] demonstrated that performance of joint attention improves owing to visual development triggered by learning progress. Other studies utilized saturation of reaching performance as a trigger for expansion of a search space [7], [8]. However, these triggers for development are task-dependent, that is, they may not synchronize with learning for other tasks and thus have difficulty in facilitating these learning. Not a task-dependent trigger but an independent one is required to explain the roles of maturational constraints in infants, who simultaneously acquire multiple abilities.

This study is partially supported by JSPS Grants-in-Aid for Specially Promoted Research (Research Project Number: 24000012).

The authors are with the Graduate School of Engineering, Osaka University. E-mail: {yuji.kawai, yukie, asada}@ams.eng.osaka-u.ac.jp

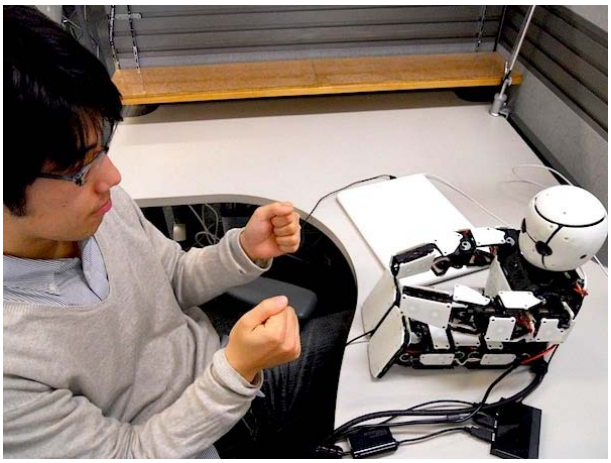


Fig. 1. A robot learns sensorimotor association through human-robot interaction, which yields the MNS.

We propose autonomous development triggered by self-organization in a perceptual space. As one of the perception, visual acuity improves when the self-organization in the visual space converges temporarily. This visual development is task-independent because perceptual self-organization can occur in learning for various cognitive functions. We investigate the role of the self-triggered development in learning for the mirror neuron system (MNS) [9], [10]. There is a hypothesis that the MNS is acquired as a by-product of sensorimotor associative learning [11], [12]. Our previous work [13] showed that immature vision enables a robot to acquire the association between self-motor commands and the same motions executed by other (i.e., a function of the MNS). A proper scheduling of visual development was crucial for the robot to experience long enough immature vision to find the correspondence. However, the visual development in the previous work was scheduled by a designer. This paper aims to show that the development triggered by self-organization in visual space would adaptively change the developmental speed, which allows a robot to acquire clearer self-other correspondence.

The rest of this paper is organized as follows: We first explain the issue and assumptions for the learning of the MNS in Section II. Section III describes a model for the MNS acquired through self-triggered visual development. Section IV demonstrates that the self-triggered development accelerates and enhances learning of the MNS. The discussion and the conclusion are given in the last section.

978-1-4673-1735-1/12/S31.00 ©2012 IEEE

II. Problem Setting and Assumptions for the Development of the MNS

A robot learns sensorimotor association between self-motor commands and observed motions through a face-to-face interaction with an experimenter (see Fig. 1). Our model makes the following assumptions:

1) The robot and the experimenter have the same repertoire of motions, which are arm gestures in the current experiment. They execute one of the gestures randomly during learning.
2) The robot's sensation is only visual perception. The robot detects the arm gestures of the experimenter and those of itself as optical flow vectors.
3) The robot's head is fixed. Thus, the robot observes only its arms while it detects the whole upper body of the experimenter.
4) The experimenter responds to the robot's motion with a certain length of time delay.

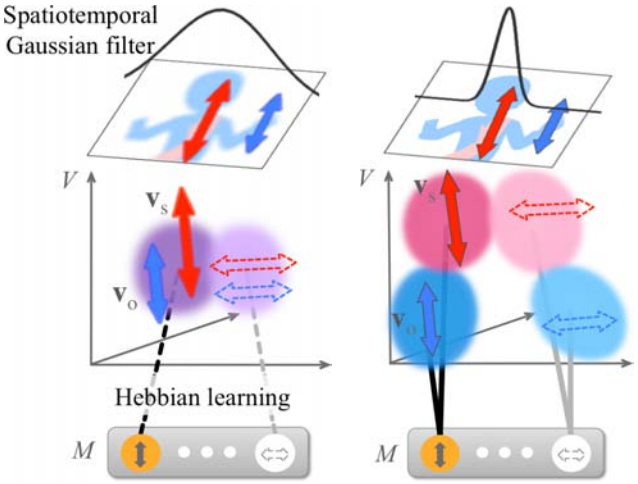
Under these assumptions, the robot learns association between self-motor commands and the corresponding other's motions, namely the main function of the MNS. Of importance in the acquisition of the MNS is finding similarity between self- and other-motions. However, the optical flow vectors observed by the robot usually have differences between them. The experimenter's motions include his head's and torso's movements as well as arm gestures (the 3rd assumption) and have delay from the robot's motions (the 4th assumption). We therefore embed visual maturational constraints in the robot so that it can inhibit these self-other differences.

III. A Model for the MNS Acquired Through Visual Development

A. Overview

Fig. 2 depicts a model for the MNS acquired through self-triggered visual development [13]. The upper layer is the visual space V, where optical flow vectors detected from the robot's vision are self-organized. The red and the blue arrows in Fig. 2 represent the self-motions ($\mathbf{v}_s$) and other-motions ($\mathbf{v}_o$), respectively. The lower layer is the motor space M consisting of a repertoire of motor commands. The followings explain the procedure at each learning step, where ψ represents a maturational clock:

1) *Visual development*: Observed optical flow vectors are smoothed by spatiotemporal Gaussian filters, whose variance is defined by using ψ. Visual acuity develops when ψ increases depending on the self-organization in V.
2) *Self-organization in V*: The filters' outputs $\mathbf{v}$ are mapped into V and then self-organized, which is shown as the ellipses in Fig. 2. No further change in the self-organization causes an increase in ψ.
3) *Associative learning*: The model learns mapping between visual clusters in V and motor commands in M by Hebbian learning.



(a) Early stage of development (b) Later stage of development

Fig. 2. A model for the emergence of the MNS acquired through visual development. Observed optical flow vectors are spatiotemporally filtered with a Gaussian. The width of the filter is determined by the progress of self-organization in V. In the early stage of development (a), the robot associates between motor commands and visual clusters, which do not differentiate self- from other-motions. In the later stage (b), a narrow filter creates the differentiated clusters while maintaining the association between other's motions with the corresponding self's motor commands, namely the MNS.

In the early stage of learning (see Fig. 2 (a)), the spatiotemporal Gaussian filters have a large variance. These blurred images prevent the robot from detecting the differences between $\mathbf{v}_s$ and $\mathbf{v}_o$. Thus the motor commands are associated with the visual clusters, which contain both $\mathbf{v}_s$ and $\mathbf{v}_o$ (purple ellipses). In the later stage of development (see Fig. 2 (b)), matured vision allows self-other differentiation for the visual clusters (i.e., red ellipses and blue ones). We assume here that the initially-acquired connections between visual clusters and motor commands are preserved through learning. Thus, the model acquires the association between self-motor commands and other's motion, i.e., the MNS.

The following sections describe the details of the three processes.

B. Mechanism of Visual Development

Fig. 3 shows the mechanism of visual development inspired by behavioral studies on infants' vision [14]–[16]. They suggested that young infants have not only spatially- [14], [15] but also temporally-immature vision [16]. It has been reported, for example, that 5-month-old infants have difficulty in discriminating their own image with 2 second of delay from that with no delay [16].

In our model, optical flow vectors as shown in Fig. 3(a) are smoothed by spatiotemporal Gaussian filters. The variance of the filters gradually decreases in order to replicate infants' visual development. Here, there are three types of Gaussian filters: directional, spatial and temporal filters.

Fig. 3(b) shows the first processing: a directional filter. We suppose optical flow vectors whose lengths and orientation are respectively $r_{t,x,y}$ and $\theta_{t,x,y}$. Here, t and (x, y) denote

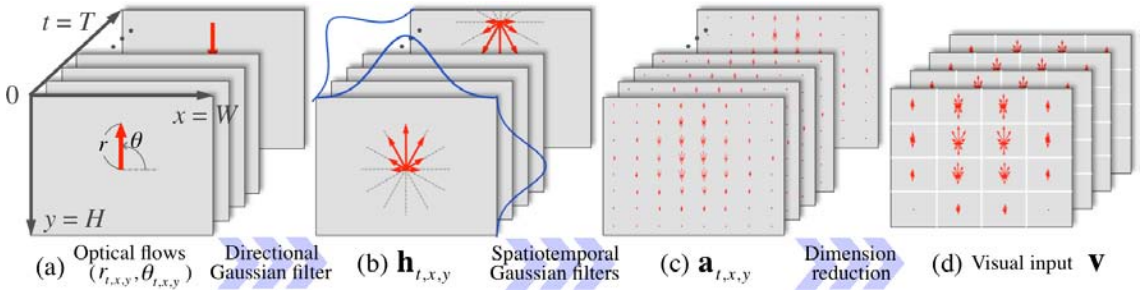


Fig. 3. Coding of optical flow vectors through spatiotemporal filters. (a) The red lines denotes optical flow vectors ($r_{t,x,y}$, $\theta_{t,x,y}$). (b) They are directionally filtered with a Gaussian to convert into histograms $\mathbf{h}_{t,x,y}$. (c) The filters' outputs $\mathbf{a}_{t,x,y}$ are given as spatiotemporally blurred $\mathbf{h}_{t,x,y}$. (d) Visual inputs $\mathbf{v}$ are the dimensionally-reduced $\mathbf{a}_{t,x,y}$.

the time and position in the image, respectively. They are converted to directional histograms $\mathbf{h}_{t,x,y}$ by a directional Gaussian filter:

$$\mathbf{h}_{t,x,y} = \big(f\left(\theta_{t,x,y}, \theta'_1, 180\right) r_{t,x,y}, \cdots, f\left(\theta_{t,x,y}, \theta'_\Theta, 180\right) r_{t,x,y}\big), \quad (1)$$

where

$$f\left(z, z', Z\right) = \exp\left(-\frac{1}{2}\left(\frac{z - z'}{Z - (Z-1)\psi}\right)^2\right). \quad (2)$$

Here, θ'_p $(p = 1, 2, \cdots, \Theta)$ represents the discretized angles, which correspond to bins of the histograms. In our model, we set $\Theta = 12$, that is, $\theta'_1 = 0°, \theta'_2 = 30°, \cdots, \theta'_\Theta = 330°$. Eq. (2) is a general definition of a Gaussian function, whose variance is determined by $\psi \in [0, 1]$. The larger ψ is, the smaller the variance of the Gaussian is.

Next, we apply spatial and temporal filters (see Fig. 3(c)): $\mathbf{h}_{t,x,y}$ is smoothed by the Gaussian filters with respect to space and time. The output of the filters $\mathbf{a}_{t,x,y}$ are given by:

$$\mathbf{a}_{t,x,y} = \sum_{t'=0}^{T} \sum_{x'=0}^{W} \sum_{y'=0}^{H} \left(f\left(t, t', T\right) \cdot f\left(x, x', W\right) \cdot f\left(y, y', H\right) \cdot \mathbf{h}_{t',x',y'}\right), \quad (3)$$

where T is the number of image frames used for observing hand gestures, and W and H are the width and height of the image, respectively. The spatiotemporal filters use the same Gaussian function f defined in Eq. (2).

Finally, a visual input $\mathbf{v}$ is calculated as the sum of $\mathbf{a}_{t,x,y}$ which is spatiotemporally discretized into four in order to reduce its dimension as shown in Fig. 3(d).

$$\mathbf{v} = \left(\sum_{t=0}^{T/4} \sum_{x=0}^{W/4} \sum_{y=0}^{H/4} \mathbf{a}_{t,x,y}, \sum_{t=0}^{T/4} \sum_{x=0}^{W/4} \sum_{y=H/4}^{H/2} \mathbf{a}_{t,x,y}, \cdots, \sum_{t=3W/4}^{T} \sum_{x=3W/4}^{W} \sum_{y=3H/4}^{H} \mathbf{a}_{t,x,y}\right) \quad (4)$$

At the beginning of learning, ψ is set to 0, and then increases gradually over learning. Fig. 4 shows sample images of $\mathbf{v}$ at (a) $\psi = 0$ and at (b) $\psi = 1$. The red lines illustrate the histograms of optical flow vectors. The left and right images show $\mathbf{v}$ detected when the robot is observing the motion of its own arm and the motion of other, respectively. Both the robot and the person were waving its/his arm horizontally. With immature vision (see Fig. 4(a)), the optical flow vectors of the both images are similar to each other. This is because the temporal smoothing inhibits the detection of the delay in other's motion. The spatial smoothing also diminishes the difference in the viewpoint, i.e., blurs the movement of other's head and torso. In the later stage of development (see Fig. 4(b)), in contrast, the differences between self- and other-motion appear, that is, the histograms of other's motion become smaller than self-motion because the matured vision enables the robot to detect the delay in other's motions. Thus the robot can differentiate self- from other-motions as ψ increases.

C. Self-organization in visual space and maturational clock

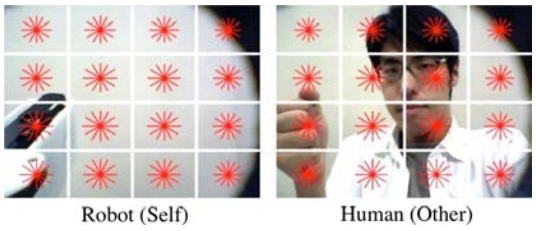
Visual inputs $\mathbf{v}$ are clustered in the visual space V. We adopt X-means algorithm [17] based on Bhattacharyya distance as a clustering method. This algorithm can determine a proper number of clusters in k-means according to Bayesian information criterion. Visual experiences cause an increase in the number of clusters, which represents progression of the visual self-organization.

We design the maturational clock ψ based on the self-organization in the visual space. The state of self-organization can indirectly represent progress of the associative learning. In order to adjust the speed of development according to the learning progress, ψ increases when the number of the clusters does not change for n steps.

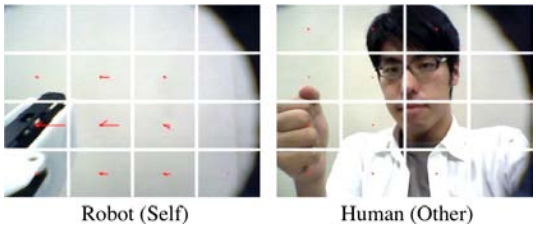
D. Visio-motor associative learning

Visual clusters are associated with motor commands by modified Hebbian learning; not only the most excited visual cluster $\mathbf{c}_{fire}$ containing the current visual input, but also its neighbors are associated with the current motor command. Let $\mathbf{c}_i$ and $\mathbf{m}_j$ be the center of the i-th cluster and the prototype vector of the j-th motor command, respectively. The connecting weight $w_{i,j}$ between $\mathbf{c}_i$ and $\mathbf{m}_j$ is updated by:

$$w_{i,j}^{t+1} = w_{i,j}^{t} + \alpha(\mathbf{c}_{fire}, \mathbf{c}_i) \cdot \beta(\mathbf{m}_j), \quad (5)$$

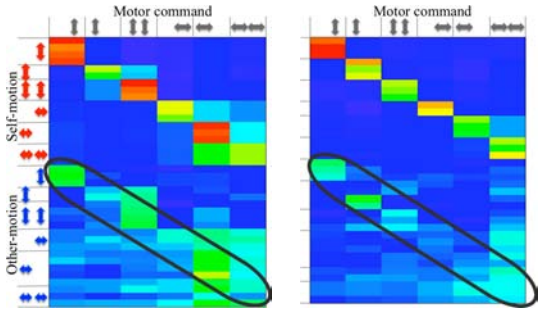


(a) Optical flow vectors coded with the most immature vision ($\psi = 0$)

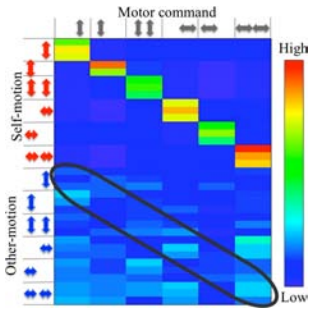


(b) Optical flow vectors coded with the most matured vision ($\psi = 1$)

Fig. 4. Samples of visual coding using spatiotemporal Gaussian filter. Immature vision (a) inhibits the detection of differences between self- (left) and other-motions (right). Matured vision (b) enables the robot to differentiate self from other.



(a) Self-triggered development ($n = 10$) (b) Time-dependent development ($\psi = t'_s$)



(c) Learning without development ($\psi = 1$)

Fig. 5. Sensorimotor mapping acquired through associative learning. The connection weights enclosed by an ellipse are correct correspondence between other's motions and self's motor commands.

where $\alpha(\mathbf{c}_{fire}, \mathbf{c}_i)$ and $\beta(\mathbf{m}_j)$ are the activities of $\mathbf{c}_i$ and $\mathbf{m}_j$, respectively:

$$\alpha(\mathbf{c}_{fire}, \mathbf{c}_i) = \exp\left(-q d_b(\mathbf{c}_{fire}, \mathbf{c}_i)^2\right) \quad (6)$$

$$\beta(\mathbf{m}_j) = \begin{cases} 1 & \text{if } \mathbf{m}_j \text{ is executed} \\ 0 & \text{else.} \end{cases} \quad (7)$$

Here q is a parameter to determine the variance of the Gaussian function of Eq. (6), and $d_b(\mathbf{c}_{fire}, \mathbf{c}_i)$ is Bhattacharyya distance between $\mathbf{c}_{fire}$ and $\mathbf{c}_i$.

Visual clusters gradually split as the model improves the visual filters and/or gains more visual experiences. The initial connecting weights of split clusters copy the neighbor clusters' one at the previous time step so that the model maintains self-other correspondence acquired with immature vision. The current connecting weight $w_{sp,j}$ between the split sp-th cluster and the j-th motor command is calculated by:

$$w^t_{sp,j} = \sum_k \left(\frac{\alpha(\mathbf{c}'_k, \mathbf{c}_{sp})}{\sum_l \alpha(\mathbf{c}'_l, \mathbf{c}_{sp})} \cdot w^{t-1}_{k,j} \right), \quad (8)$$

where $\mathbf{c}'_k$ is the center of the k-th cluster at $t-1$.

IV. Experiments

A. Experimental Setting

We evaluated the validity of self-triggered visual development on the acquisition of the MNS. We conducted the experiment using an infant-like humanoid robot as shown in Fig. 1. The robot, called M3-Neony [18], has 22 degrees of freedom and two CMOS USB cameras (640 × 480 pixels and about 30 fps) embedded in the eyes. Both the robot and an experimenter had 6 types of actions: waving the right arm, the left arm, and both arms vertically and horizontally. They executed arm gestures, in which the experimenter's response had 2 to 3 seconds of delay from the robot's motion. It has been reported that caregivers often imitate infants' motions (e.g., [19]). We however assume extremely less imitative situation: the robot and the human choose a motion *randomly* at each step in this experiment.

120 sets of visual data were acquired in advance, and then off-line learning was conducted. We set q in Eq. (6) to 10^7 based on preliminary experiments.

Maturational clock ψ had twenty one stages between 0 and 1. It moved to the next stage when the number of clusters did not change for n steps, which was set to 2, 5, 10, or 20 in the current experiment. The learning then finished n steps after ψ became 1. We compared this self-triggered development with the one depending on the learning time step t_s. The comparative models defined ψ as $\psi = t'_s$, $\sqrt{t'_s}$, t'^2_s, or t'^3_s ($t'_s = t_s/300$).

B. Effect of self-triggered visual development

Fig. 5 depicts the connecting weights between self-motor commands and visual clusters after learning: (a) with self-triggered development ($n = 10$), (b) with time-dependent development ($\psi = t'_s$), and (c) without development ($\psi = 1$). The rows and columns correspond to the visual clusters (self-motions in the upper side and other-motions in the lower side) and motor commands, respectively. The arrows on the left and on the top denote the gestures. For example, the leftmost column shows the vertical movement of the right hand, the second leftmost the left hand, and so on.

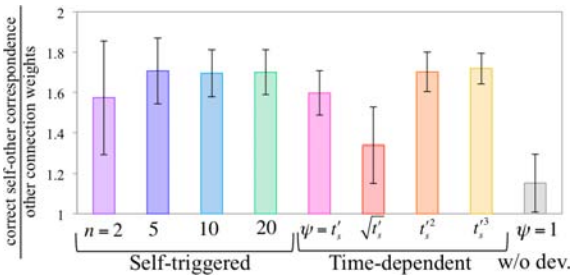


Fig. 6. Connection weights between self-motor commands and the corresponding other-motions

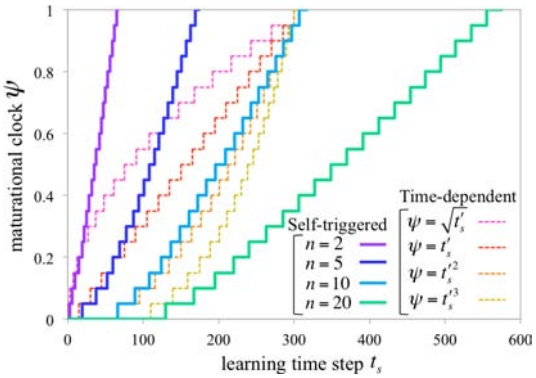


Fig. 7. Schedules of visual development. The solid and the dashed lines are self-triggered and time-dependent development respectively.

Each result in Fig. 5 indicates the strong connecting weights from the top-left corner to the middle-right. These connecting weights are between the motor commands and the corresponding motions produced by self. It is easier to acquire the association concerned with self-motions because the robot's motions are highly contingent. On the other hand, learning the proper association with other's motions is difficult because they do not always coincide with the self-motions. Fig. 5 shows that only learning with visual development, especially with (a) self-triggered development, results in the strong connections between self-motor commands and the corresponding other-motions, which are enclosed by the ellipses. These connections are the correct correspondence between self- and other-motions, namely the MNS.

The strength of self-other correspondence acquired in each condition is summarized in Fig. 6. The vertical axis shows the ratio of the mean of connecting weights of the MNS to the mean of the others. These results are the averages of twenty trials. The higher value indicates the clearer acquisition of the correct self-other correspondence. As shown in Fig. 6, the self-triggered development with $n = 5$ or more yields stronger self-other correspondence. The time-dependent development with $\psi = t_s^{'2}$ and $t_s^{'3}$ also results in higher performance than with $\psi = \sqrt{t_s'}$ and t_s'.

Fig. 7 shows the developmental schedule in each condition to investigate the cause of the above result. The solid and the dashed lines are self-triggered and time-dependent development, respectively. The maturational clock defined by $\psi = t_s^{'2}$ and $t_s^{'3}$, which enables the acquisition of the self-other correspondence, increases slowly at first and then progresses rapidly. The self-triggered development with $n = 5$ or more shows comparable slower increase in the early stage of learning. This finding suggests that self-triggered development and properly defined time-dependent development can enhance the emergence of the MNS. Futhermore, the self-triggered development enables a quicker and clearer acquisition of the MNS than other conditions.

An appropriate n might depend on the number of the motor repertoires. We used 6 types of motion in the current experiment. We conjecture that n close to the number of the motor repertoires may achieve higher performance, which needs to be further investigated.

Novel visual inputs in the early stage of development facilitate the generation of new visual clusters. Increasing number of clusters decelerates progression of maturational clock. The period of immature vision allows the robot to associate other's motion and self's motor commands because the immature vision diminishes the self-other differences. In the later stage of development, saturation of the number of visual clusters improves the robot's vision rapidly. The robot consequently finishes the learning while maintaining the initially-acquired association. The adaptive change of developmental speed (i.e., slower in the early stage and faster in the later stage) enables the acquisition of clearer correspondence between self and other.

C. Activation of the MNS after learning

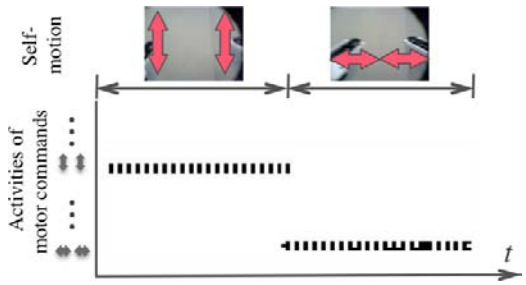
The acquired connecting weights enable the robot to activate motor commands corresponding to visual inputs. We tested activation of motor commands during observation of self- and other-gestures after sensorimotor learning. This experimental paradigm is inspired by neuroscientific studies on mirror neurons (e.g., [20], [21]), which recorded discharge of a mirror neuron in monkeys.

The activity of the j-th motor command $\beta_i'(\mathbf{m}_j)$, which is associated with the i-th visual cluster, is calculated by:

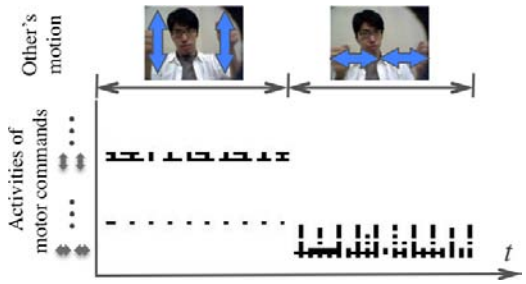
$$\beta_i'(\mathbf{m}_j) = \begin{cases} 1 & \text{if } w_{i,j} \cdot \alpha(\mathbf{v}, \mathbf{c}_i) \geq T_h \\ 0 & \text{else,} \end{cases} \quad (9)$$

where $\mathbf{v}$ is a visual input, and T_h is the threshold of activities, which was set to 8. The parameter q in Eq. (6) was set to 1. We used the connecting weights obtained by the learning through self-triggered development (see Fig. 5(a)).

Fig. 8 illustrates the activities of motor commands when the robot executed arm gestures (a) and when the robot observed other's motion (b). The rows correspond to the columns in Fig. 5. The black bars show the raster plot of the activation of the motor commands, i.e., $\beta_i'(\mathbf{m}_j) = 1$. The robot and the experimenter moved both of its/his hands vertically and then horizontally. The robot's observation successfully activated the corresponding motor commands regardless of self- or other-motions. These activities of motor commands enable the robot to understand other's motions based on self-motions and to imitate them. This result is similar to responses of monkey's mirror neurons [20], [21].



(a) The robot's execution of hand gesture



(b) Observation of other-motions

Fig. 8. Activation of the MNS during the robot's execution of gesture (a) and observation of other's gesture (b). Both the robot and the experimenter were moving both hands first vertically and switched to horizontally. The robot successfully activates the corresponding motor commands when observing other's motions, which is the property of the MNS.

V. Discussion and conclusion

We have suggested that perceptual development triggered by its self-organization facilitates cognitive learning. In our experiment, self-triggered visual development was applied to learning for the MNS. The development maintained a long enough period of immature vision which could inhibit the differences between self- and other-motions. The association acquired in this period resulted in the MNS. In comparison to the time-dependent development, the proposed method can adaptively change the developmental speed, which achieve quicker acquisition of the clearer MNS.

The self-triggered development may produce greater effect when considering robots' motor development. Given gradual increase in motor repertoires during learning, the perceptual development should adapt to it. In our model, novel visual experiences caused by motor development would create new visual clusters, which temporally decelerates the perceptual development. Such developmental adaptation may yield the chances to associate new motor repertoires with non-differentiated visual clusters. On the other hand, the perceptual development pre-defined by a designer have difficulty in adapting to the motor development.

Infants acquire simultaneously various cognitive abilities, which may be facilitated by their immature perception. The self-organization in perceptual space is important for such cognitive functions. We suggest that the self-organization is a task-independent and reasonable index to indirectly measure the task performance. For example, joint attention, which is a process to look where other is looking, is acquired in infancy [22]. A robot needs to learn a mapping between observation of other's face and motor commands to achieve joint attention [6]. In the visual space for joint attention, the clusters may represent directions of other's gaze. Such organization in the visual space would affect the task accuracy of joint attention. The previous works [6]–[8] showed that maturational constraints lifted by the task performance make the learning more efficient. Thus we believe that perceptual development triggered by its self-organization achieves as effective learning as in the previous works through the trigger is task-independent. We attempt to evaluate the proposed self-triggered development in other multiple tasks.

References

[1] D. Bjorklund, "The role of immaturity in human development," *Psychological Bulletin*, vol. 122, no. 2, pp. 153–169, 1997.

[2] G. Turkewitz and P. Kenny, "The role of developmental limitations of sensory input on sensory/perceptual organization," *Journal of Developmental and Behavioral Pediatrics*, vol. 6, no. 5, pp. 302–306, 1985.

[3] E. L. Newport, "Maturational Constraints on Language Learning," *Cognitive Science*, vol. 14, no. 1, pp. 11–28, 1990.

[4] J. L. Elman, "Learning and development in neural networks: the importance of starting small," *Cognition*, vol. 48, no. 1, pp. 71–99, 1993.

[5] M. Dominguez and R. A. Jacobs, "Developmental constraints aid the acquisition of binocular disparity sensitivities," *Neural Computation*, vol. 15, no. 1, pp. 161–182, 2003.

[6] Y. Nagai *et al.*, "Learning for joint attention helped by functional development," *Advanced Robotics*, vol. 20, no. 10, pp. 1165–1181, 2006.

[7] A. Baranes and P. Oudeyer, "The interaction of maturational constraints and intrinsic motivations in active motor development," in *Proc. of the 1st Joint Intl. Conf. on Development and Learning and on Epigenetic Robotics*, 2011.

[8] M. Lee *et al.*, "Staged competence learning in developmental robotics," *Adaptive Behavior*, vol. 15, no. 3, pp. 241–255, 2007.

[9] G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annu. Rev. Neurosci.*, vol. 27, pp. 169–192, 2004.

[10] G. Rizzolatti and C. Sinigaglia, *Mirrors in the brain: how our minds share actions, emotions*. Oxford University Press, USA, 2008.

[11] C. Catmur *et al.*, "Associative sequence learning: the role of experience in the development of imitation and the mirror system," *Philosophical Transactions of the Royal Society B*, vol. 364, no. 1528, pp. 2369–2380, 2009.

[12] E. Ray and C. Heyes, "Imitation in infancy: The wealth of the stimulus," *Developmental Science*, pp. 92–105, 2011.

[13] Y. Nagai *et al.*, "Emergence of mirror neuron system: Immature vision leads to self-other correspondence," in *Proc. of the 1st Joint Intl. Conf. on Development and Learning and on Epigenetic Robotics*, 2011.

[14] R. Wilson, "Development of spatiotemporal mechanisms in infant vision," *Vision Research*, vol. 28, no. 5, pp. 611–628, 1988.

[15] T. Banton *et al.*, "Infant direction discrimination thresholds," *Vision Research*, vol. 41, no. 8, pp. 1049–1056, 2001.

[16] K. Hiraki, "Detecting contingency: A key to understanding development of self and social cognition," *Japanese Psychological Research*, vol. 48, no. 3, pp. 204–212, 2006.

[17] D. Pelleg and A. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," in *Proc. of the 17th Intl. Conf. on Machine Learning*, 2000, p. 727.

[18] H. Ishiguro *et al.*, "Humanoid platforms for cognitive developmental robotics," *International Journal of Humanoid Robotics*, vol. 8, no. 3, pp. 391–418, 2011.

[19] G. Moran *et al.*, "Patterns of maternal and infant imitation during play," *Infant Behavior and Development*, vol. 10, no. 4, pp. 477–491, 1987.

[20] G. Rizzolatti *et al.*, "Premotor cortex and the recognition of motor actions," *Cognitive brain research*, vol. 3, no. 2, pp. 131–141, 1996.

[21] M. Umilta *et al.*, "I Know What You Are Doing: A Neurophysiological Study," *Neuron*, vol. 31, no. 1, pp. 155–165, 2001.

[22] G. Butterworth and N. Jarrett, "What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy," *British journal of developmental psychology*, vol. 9, no. 1, pp. 55–72, 1991.