Emergence of Mirror Neuron System: Immature vision leads to self-other correspondence

Yukie Nagai*, Yuji Kawai*, and Minoru Asada*†
*Graduate School of Engineering, Osaka University
†JST ERATO Asada Synergistic Intelligence Project
2-1 Yamada-oka, Suita, Osaka, 565-0871 Japan
Email: {yukie, yuji.kawai, asada}@ams.eng.osaka-u.ac.jp

***Abstract*—The question of how the mirror neuron system (MNS) develops has attracted increased attention of researchers. Among various hypotheses, a widely accepted model is associative sequence learning, which acquires the MNS as a by-product of sensorimotor learning. The model, however, cannot discriminate self from others since it adopts too much simplified sensory representations. We propose a computational model for early development of the MNS, which is originated in immature vision. The model gradually increases the spatiotemporal resolution of a robot's vision while the robot learns sensorimotor mapping through primal interactions with others. In the early stage of development, the robot interprets all observed actions as equivalent due to a lower resolution, and thus associates the non-differentiated observation with motor commands. As vision develops, the robot starts discriminating actions generated by self from those by others. The initially acquired association is, however, maintained through development, which results in two types of associations: one is between motor commands and self-observation and the other between motor commands and other-observation (i.e., what the MNS does). Our experiments demonstrate that the model achieves early development of the MNS, which enables a robot to imitate others' actions.**

I. INTRODUCTION

The mirror neuron system (MNS) discharges both when executing a specific action and when observing the same action by other individuals [1], [2]. Many researchers in neuroscience, developmental psychology, and even robotics have been investigating properties of the MNS, e.g., how it represents self-other correspondence and how it leads to social behaviors. Notable findings are the ability to understand others' actions (e.g., [3], [4]) and imitation (e.g., [5], [6]). However, despite many findings, the origin of the MNS remains a mystery.

The question of whether the MNS is nature or nurture is still controversial. We support a hypothesis that the MNS develops by postnatal learning since human ability to imitate varies depending on experiences [7]. Heyes and colleagues [7], [8] proposed an associative sequence learning (ASL) model, which hypothesizes that infants acquire the MNS as a by-product of sensorimotor learning. The model consists of sensory representations and motor ones, which are weakly and unsystematically connected before learning. Infants are supposed to learn the sensorimotor mapping using the Hebb's rule. An important assumption here is that infants are often imitated by others, especially by caregivers, which enables infants to detect correspondence between self and other. Many researchers have suggested that experiences of being imitated is crucial in the development of the MNS [7]–[10]. The ASL model, however, cannot determine whether an observed action is produced by self or other. Too much simplified sensory representations prevent the model from detecting temporal delay in other's actions, different perspective between self and other, and so on, which are important in discriminating self from others.

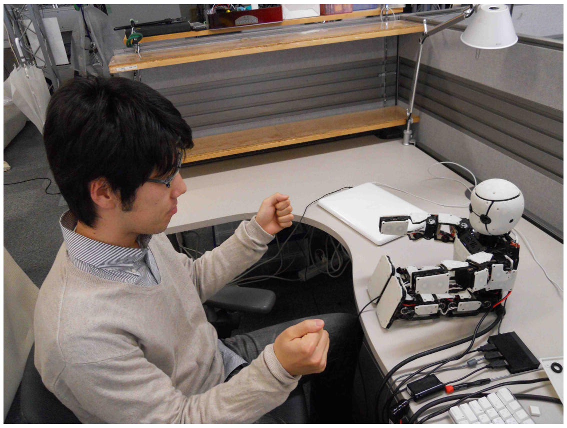


Fig. 1. Human-robot interaction for development of the MNS

Neither previous computational models of the MNS nor those of imitation take self-other discrimination into account. Kuniyoshi et al. [11] showed that equivalence between optical flows detected from motion of self and those of other serves as the ability of imitation. Sensorimotor learning only through observation of self enabled their robot to "imitate" equivalent other's actions. Chaminade et al. [12] also employed the idea of equivalence. Their robotic hand "imitated" the posture of other's hand based on close similarity between visual appearance of self and other. In contrast, the models proposed by Asada et al. [13] and Minato et al. [14] differentiate self from others. Their models make a robot/an infant model first observe motions of self and other separately, map the observed motions into different perceptual spaces, and then associate them with self-motor commands. However, a remaining question is how a robot/an infant model can determine which observed motion is self or other. Such differentiation

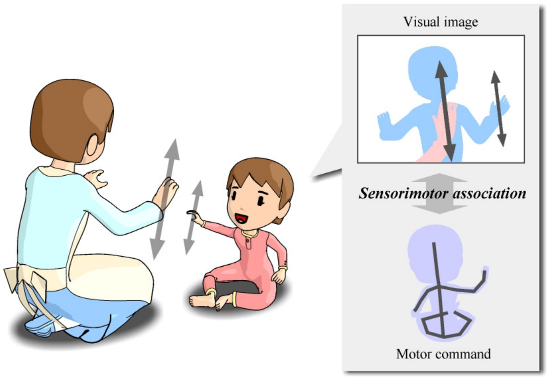


Fig. 2. An infant learns sensorimotor mapping through interactions with a caregiver, which produces the MNS as a by-product of associative learning.

is necessary for further social development such as contingent interaction and turn taking. It has even been suggested that infants are born with a vague or no border between self and others [15]. A developmental model of the MNS has to explain how infants acquire self-other correspondence while learning to differentiate self from others.

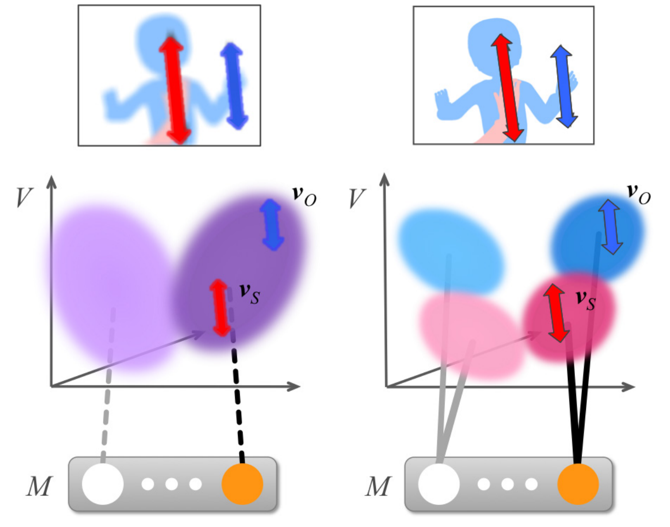
We propose a computational model for early development of the MNS originated in immature vision. A robot is supposed to learn sensorimotor mapping through primal interactions with others (see Fig. 1). Meanwhile, the robot's vision develops as it gains more experiences. In the early stage of development, the robot observes motions of self and those of other only in a lower spatiotemporal resolution. Such immature vision prevents the robot from detecting differences between self and other (e.g., different length of delay in actions and different appearances of them), and thus makes the robot associate between non-differentiated observations and motor commands. Of importance here is that the non-differentiated perception leads to correspondence between self and other in the latter stage of development. As vision develops, the robot starts discriminating self from other while maintaining the initially-acquired association, that is, the mapping both between self-motions and motor commands and between other-motions and motor ones (i.e., what the MNS does). Our model thus enables a robot to acquire both self-other correspondence and discrimination simultaneously through visual development.

The rest of this paper is organized as follows: Section II defines the issue we address and assumptions. Our proposed model for emergence of the MNS is described in Section III. Section IV then demonstrates how the model enables a robot to acquire the MNS and to imitate others' actions. Experimental results are discussed in Section V with concluding remarks.

II. PROBLEM SETTING AND ASSUMPTIONS

We assume a face-to-face interaction between an infant (or a robot) and a caregiver as shown in Fig. 2. The infant learns association between self-motor commands and observed body motions under the following assumptions:

1) Both the infant and the caregiver have the same repertoire of motions (hand gestures in our experiment) and execute them one by one during sensorimotor learning.
2) The infant only uses visual perception in the current experiment. The hand gestures of the infant and the caregiver are detected as optical flows.
3) The infant has his own perspective and limited view, that is, he can see only his arms while observing the whole upper body of the caregiver.
4) The caregiver responds to the infant's motions with some delay. Her responding gestures are sometimes the copy of the infant's ones.



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

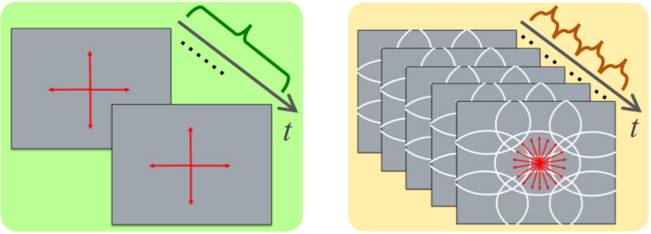
Fig. 3. A model for emergence of the MNS originated in immature vision. In the early stage of development (a), motor commands are roughly associated with visual clusters, which do not yet discriminate self from others. In the latter stage (b), clusters become differentiated while maintaining the association with the same motor commands (i.e., the MNS).

Under these assumptions, we address two issues related to the MNS: The first one is *self-other differentiation*. The infant learns to discriminate self from others by detecting differences in observed motions. Differences are, for example, caused by the infant's specific perspective (the 3rd assumption) and the temporal delay in the caregiver's motions (the 4th). The second issue concerns *self-other correspondence*. The infant learns to associate between self-motor commands and corresponding caregiver's motions as well as between motor commands and observed self's motions.

III. A MODEL FOR EMERGENCE OF MNS

A. Basic idea

Fig. 3 illustrates the proposed model for emergence of the MNS: (a) the early stage and (b) the latter stage of development. The model consists of two layers: The upper layer is the visual space V, which maps optical flows detected from an infant's vision. The red and the blue arrows denote the motion of self and other, respectively. The lower layer is the motor space M, which contains the repertoire of motor commands.



(a) Immature vision with 1 spatial, 4 directional, and 1 temporal samplings

(b) Matured vision with 9 spatial, 18 directional, and 4 temporal samplings

Fig. 4. Mechanism of visual development. Out of three developmental stages, the first (a) and last stages (b) are depicted.

An infant learns visuomotor mapping while he develops visual perception. In the early stage of development (see Fig. 3 (a)), the infant perceives optical flows only in a lower spatiotemporal resolution. Thus, the infant recognizes the caregiver's motions as equivalent to the corresponding self-motions, which forms non-discriminated clusters in V (i.e., purple ellipses). The non-discriminated clusters then establish rough association with motor commands in M by Hebbian learning. In the latter stage of development (see Fig. 3 (b)), the infant comes to perceive visual inputs with a higher resolution. Clusters of self-motions and other-motions gradually become discriminated (i.e., red ellipses and blue ones, respectively). Of importance here is that the initially-acquired association is maintained through development. The mapping that each motor command had with a non-differentiated cluster is now established both with self-motion and other-motion. That is, the properties of the MNS (i.e., self-other differentiation and correspondence) are acquired through associative learning.

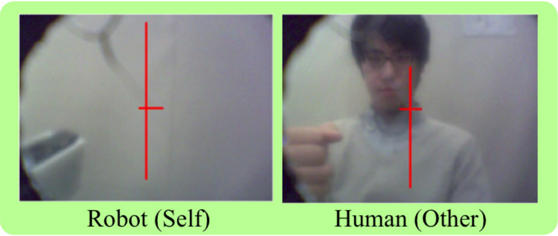
The following sections describe the mechanism of visual development (Section III-B), the method for clustering visual inputs and associative learning (III-C), and that of imitation after learning (III-D).

B. Mechanism of visual development

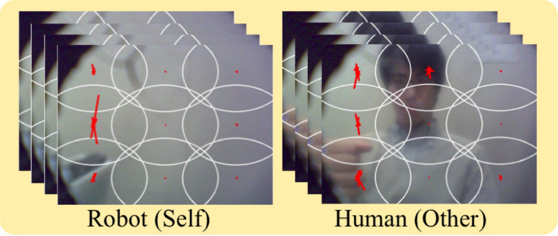
We replicate infants' visual development by changing the spatiotemporal resolution of a robot's vision. Fig. 4 illustrates how to code optical flows detected from motions of self and other. Out of three developmental stages, the first (a) and last stages (b) are depicted here. There are three types of resolutions to develop:

1) Spatial resolution: The first one concerns receptive fields denoted by white circles in Fig. 4 (b). The number of receptive fields increases as vision develops: It starts with 1 receptive field covering a whole image (see Fig. 4 (a)), changes into 4 fields dividing an image into 2 × 2, and ends with 9 fields dividing into 3 × 3 (see Fig. 4 (b)). Optical flows detected in a robot's vision are accumulated and then coded as a histogram in each receptive field. The mechanism of employing histograms except the developmental process was inspired by [11], [16].

2) Directional resolution: The second one is directional resolution to discretize optical flows (see red arrows in Figs. 4 (a) and (b)). A robot starts only with 4 directional samplings, that is, optical flows are discretized every 90 deg as shown in Fig. 4 (a). The number of samplings then increases to 8 (i.e., every 45 deg) and finally to 18 (i.e., every 20 deg as shown in Fig. 4 (b)) as vision develops.



(a) Optical flows coded with immature vision. *No* significant difference between self and other is observed due to a lower spatiotemporal resolution.



(b) Optical flows coded with matured vision. Differences between self and other become visible due to a higher resolution.

Fig. 5. Sample images showing optical flows detected from self-motion (left) and other-motion (right). The developmental stages of (a) and (b) correspond to those in Figs. 4 (a) and (b), respectively.

3) Temporal resolution: The third one is temporal resolution, which defines the length of a time window to accumulate optical flows (see "}" at the upper-right in Figs. 4 (a) and (b)). A robot first has only a single longer temporal window, which corresponds to the duration of a single gesture (see Fig. 4 (a)). The temporal resolution then increases to 2 and finally to 4 as shown in Fig. 4 (b), which gradually enables a robot to detect temporal delay in other's actions.

Let N_s, N_d, and N_t be the number of spatial, directional, and temporal samplings, respectively. Visual inputs $\mathbf{v}$ are defined by:

$$\mathbf{v} = [[\ \mathbf{h}_{1,1}\ \ \ldots\ \ \mathbf{h}_{N_s,1}\ ] \ldots [\ \mathbf{h}_{1,N_t}\ \ \ldots\ \ \mathbf{h}_{N_s,N_t}\ ]]^T \qquad (1)$$

$$\text{where}\ \ \mathbf{h} = [\ v_1\ \ \ldots\ \ v_{N_d}\ ]^T \qquad (2)$$

$\mathbf{h}$ is the histogram detected in each receptive field, and v is the length of each flow vector with a directional selectivity. The number of dimensions of $\mathbf{v}$ thus increases as vision develops (i.e., 4 = 1 × 4 × 1 in the first stage, 64 = 4 × 8 × 2 in the second, and 648 = 9 × 18 × 4 in the last). All the three types of development are simultaneously triggered by increases in visual experiences.

Fig. 5 gives sample images (a) with immature vision and (b) with matured vision. The left and right images show optical flows detected when a robot was observing the motion of its own hand (the robot's left hand appears at the left lower corner of the image) and the motion of other, respectively. Both the robot and the person were moving its/his hand up and down. With immature vision (see Fig. 5 (a)), the robot does not detect significant difference between self and other due to a lower spatiotemporal resolution. In contrast, the robot with matured vision (see Fig. 5 (b)) detects differences in the spatial and directional properties of optical flows. Difference in the temporal delay is another crucial cue in discriminating self from other though it is not shown in Fig. 5 (b).

C. Clustering of visual inputs and sensorimotor learning

Visual inputs $\mathbf{v}$ are clustered in the visual space V (the upper layer in Fig. 3). We apply X-means algorithm [17] using Bhattacharyya distance because the algorithm can automatically determine an appropriate number of clusters. The process of increasing clusters shows how self and other are gradually discriminated. Refer to [17] for the detailed mechanism.

Our model then learns association between visual clusters and motor commands. We apply a modified Hebbian rule, which connects an activated motor command not only with the most excited visual cluster, which contains the current visual input, but also with its neighbors depending on the distance. Clusters which are too far from the activated one are inhibited to facilitate learning. Let $\mathbf{v}_i$ $(i = 1, 2, \cdots, N_v)$ and $\mathbf{m}_j$ $(j = 1, 2, \cdots, N_m)$ be the prototype vectors of visual clusters and motor commands, respectively. The connecting weights $w_{i,j}$ between $\mathbf{v}_i$ and $\mathbf{m}_j$ are updated by:

$$w_{i,j}(t+1) \;=\; w_{i,j}(t) + \alpha(\mathbf{v}_i) \cdot \beta(\mathbf{m}_j), \tag{3}$$

where $\alpha(\mathbf{v}_i)$ and $\beta(\mathbf{m}_j)$ are the activities of $\mathbf{v}_i$ and $\mathbf{m}_j$ calculated by:

$$\begin{aligned} \alpha(\mathbf{v}_i) \;=\; & a \exp\left(-a\pi d_b(\mathbf{v}_{fire}, \mathbf{v}_i)^2\right) \\ & -(a-1)\exp\left(-(a-1)\pi d_b(\mathbf{v}_{fire}, \mathbf{v}_i)^2\right) \end{aligned} \tag{4}$$

$$\beta(\mathbf{m}_j) = \begin{cases} 1 & \text{if } \mathbf{m}_j \text{ is executed} \\ 0 & \text{else.} \end{cases} \tag{5}$$

$\mathbf{v}_{fire}$ is the most excited visual cluster, a is the parameter to determine the sharpness of a Gaussian function, and $d_b(\mathbf{x}, \mathbf{y})$ is Bhattacharyya distance between $\mathbf{x}$ and $\mathbf{y}$. $\alpha(\mathbf{v}_i)$ thus works as a DOG (difference of two Gaussians) filter to represent lateral inhibition of Hebbian learning.

Note that visual clustering and sensorimotor learning are conducted iteratively. Visual clusters gradually separate as the model increases spatiotemporal resolution and/or gains more visual experiences. If the i-th cluster is divided into the i_1-th and i_2-th clusters, the connecting weight with the j-th motor command is copied by:

$$w_{i_1,j}(t) = w_{i_2,j}(t) = w_{i,j}(t), \tag{6}$$

so that the model maintains self-other correspondence acquired with immature vision.

D. Imitation after sensorimotor learning

The acquired sensorimotor mapping can be used for imitating other's actions. Let $\mathbf{v}_o$ be a new visual input when a robot observes other's motion. The motor command $\mathbf{m}$ to execute is calculated by:

$$\mathbf{m} = \sum_j \left(\left(\beta'(\mathbf{m}_j) / \sum_k \beta'(\mathbf{m}_k) \right) \cdot \mathbf{m}_j \right), \tag{7}$$

where $\beta'(\mathbf{m}_j)$ is the activity of the j-th motor command determined by:

$$\begin{aligned} \beta'(\mathbf{m}_j) \;=\; & \sum_i w_{i,j} \left\{ a \exp\left(-a\pi d_b(\mathbf{v}_o, \mathbf{v}_i)^2\right) \right. \\ & \left. -(a-1)\exp\left(-(a-1)\pi d_b(\mathbf{v}_o, \mathbf{v}_i)^2\right) \right\}. \end{aligned} \tag{8}$$

That is, the robot imitates other's motions by integrating motor commands with a weight proportional to their activity. This mechanism thus enables interpolation of novel actions.

IV. EXPERIMENTS

A. Experimental setting

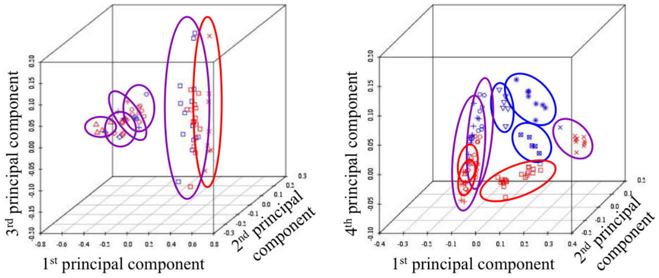
We evaluated the proposed model using an infant-like humanoid robot as shown in Fig. 1. The robot, called M3-Neony [18], has 22 degrees of freedoms (DoFs) and two CMOS USB cameras (640 × 480 pixels) embedded in the eyes. It was equipped with six types of hand gestures: waving the right hand, the left hand, or the both hands vertically or horizontally using 6 DoFs in its arms. An experimenter, who had the same repertoire of gestures, responded to the robot's actions with 30% of imitation and 70% of randomly-selected gestures. The responding gestures always had 2 to 3 seconds of delay from the robot's ones.

Learning experiments were conducted off-line. The robot simulated interactions by iteratively using 120 visual data obtained beforehand. The three-staged visual development was triggered when the robot received first 70 types of data (from the first to second stages) and then 105 data (from the second to third). These numbers as well as the parameter a in Eqs. (4) and (8), which was set at 20, were determined from preliminary experiments.

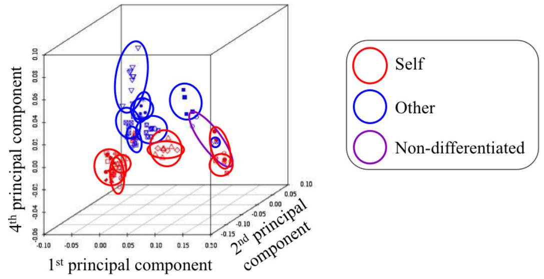
B. Self-other differentiation through visual development

Fig. 6 depicts the gradual separation of clusters for self- and other-motion in (a) the first stage, (b) the second stage, and (c) the last stage of development. We applied principal component analysis just to visualize clusters. Red points and blue ones denote an individual visual input corresponding to motion of self and other, respectively, which the robot did not know before associative learning. Ellipsoids enclosing the points indicate clusters the robot acquired through visual development (self-motion in red, other in blue, and non-discriminated in purple).

In the early stage of development, most of the clusters contained both self- and other-motions as shown in Fig. 6 (a). The robot could not differentiate between self and other due to a lower visual resolution. Once the resolution improved, the robot started differentiating self from other. Some of the clusters in Fig. 6 (b) already contained either self- or other-motions, and finally, the robot achieved clear separation in



(a) First stage of development (b) Second stage of development



(c) Last stage of development

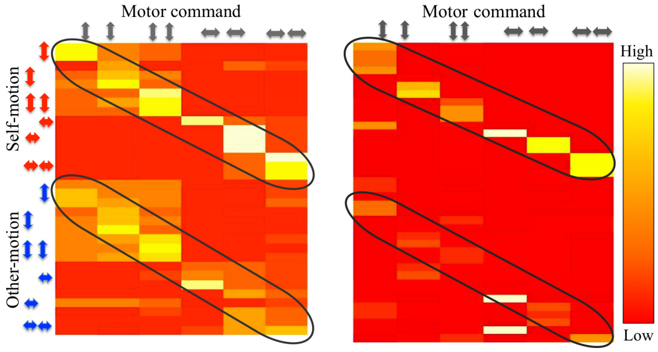
Fig. 6. Discrimination between self and other through visual development. Non-discriminated clusters (purple ellipses) in (a) gradually split into clusters for self (red ellipses) and other (blue ones) in (b) and then in (c).

Fig. 6 (c). This result further indicates that the robot can recognize whether a novel motions are generated by self or other. The vertical axis in Fig. 6 (c) clearly separates clusters for motions produced by self (the lower side) from those by other (the upper side). The visual mechanism to detect different appearance of other's motions and their temporal delay enabled the robot to clearly discriminate self from other.

C. Self-other correspondence originated in immature vision

The second experiment examined the role of visual development in learning self-other correspondence. Fig. 7 shows the connecting weights between self-motor commands and observed motions acquired (a) *with* visual development and (b) *without* development (i.e., the robot used only visual inputs with a higher resolution over learning). The rows correspond to the clusters of observed motions (self-motions in the upper side and other-motions in the lower side) whereas the columns motor commands. The small arrows on the left and on the top indicate the corresponding gestures. For example, the leftmost column shows the vertical movement of the right hand, the second leftmost the left hand, and so on.

First, both results in Figs. 7 (a) and (b) indicate that the correspondence between motor commands and observed motions produced by self are acquired regardless of visual development. The stronger weights from the top-left corner to the middle-right show proper association between them. It is obvious to acquire such association because the robot's motions were highly contingent. In contrast, the association between motor commands and motions of other presents a big difference. Only the model *with* visual development shown in Fig. 7 (a) acquired self-other correspondence (i.e., the stronger weights in the lower diagonal area). Associative learning during the early stage of visual development contributed to the acquisition of correspondence between self and other.



(a) Learning *with* visual development (b) Learning *without* development

Fig. 7. Sensorimotor mapping acquired through associative learning (a) *with* or (b) *without* visual development. Only (a) acquired self-other correspondence depicted as the stronger weights in the lower diagonal area of the map.

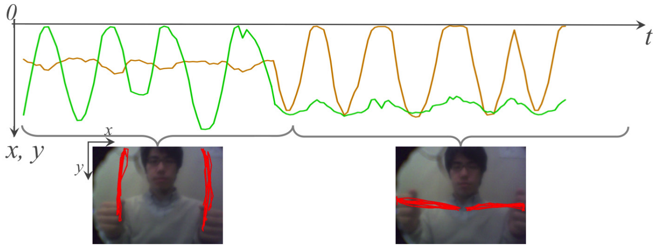
D. Imitation using acquired sensorimotor mapping

We assessed the robot's ability to imitate other's actions using the acquired self-other correspondence. Figs. 8 (a) and (b) plot the trajectories of the experimenter's right hand and the robot's left hand imitating the experimenter, respectively. The brown lines and the green ones are the horizontal and vertical positions of their hand in the image. The experimenter was moving both arms first up and down and suddenly switched to left and right. Camera images captured during each motion are shown under the graph.

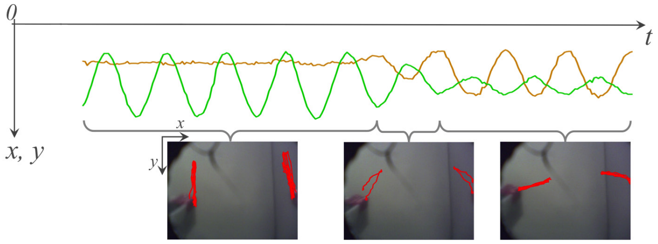
The result demonstrates qualitatively successful imitation by the robot. The robot properly selected the arms to move (i.e., the both arms) and the motion direction (i.e., first vertically and then horizontally). The delay in the robot's motion was caused by accumulating optical flows so that the robot could strongly enough activate the corresponding motor command. The gradual shift in motion direction also shows the effect of accumulation and the way of generating motor commands using Eqs. (7) and (8). Integrating motor commands with a weight proportional to their activity enabled the robot to interpolate the two different motions.

V. DISCUSSION AND CONCLUSION

How the MNS develops is still an open question. We have addressed this issue by extending the hypothesis of associative learning. Our key idea is that immature perception leads to emergence of the MNS. Visual perception with a lower spatiotemporal resolution makes infants/robots recognize motions produced by other as equivalent to those by self. Immature perception plays an important role in detecting correspondence between self and other. The experimental results showed that our model acquired the abilities of self-other discrimination as well as correspondence and enabled a robot to imitate other's actions using the MNS.



(a) Motion trajectory of an experimenter's hands



(b) Trajectory of imitative response by the robot

Fig. 8. Imitation of an experimenter's hand movement using the acquired sensorimotor mapping. The brown and the green lines indicate the horizontal and vertical position of the experimenter's right hand (in (a)) and the robot's left hand (in (b)), respectively. The robot successfully imitated the experimenter's motion in terms of the hands used in the gesture and the motion directions.

There are behavioral and neuronal evidences which support our mechanism of visual development: Atkinson et al. [19] and Wilson et al. [20] suggest that the response characteristics of receptive fields change from a low pass to a band pass type during infancy. Infants gradually come to receive a sharper image, corresponding to a higher spatial resolution. Banton et al. [21] showed a developmental change in infants' ability to discriminate motion direction. 18-weeks-old infants responded to 17 deg of stimuli while 12-weeks-olds could only 22 deg. Hiraki [22] found out that 5-month-olds have difficulty in discriminating their own image with 2 second of delay from that with *no* delay. This finding suggests that younger infants might have a lower temporal resolution.

Though our hypothesis that the MNS develops simultaneously with visual perception is not yet verified, our key idea of starting with maturational constraints can be applied to other cognitive development. Newport [23] and Elman [24] suggested that a developmental change in the capacity of infants' memory facilitates their language learning. Nagai et al. [25] demonstrated that learning for joint attention can be improved by visual development. There are even more studies supporting the importance of maturational constraints (e.g., [26], [27]), which strengthen the plausibility of our model.

For future issues, our model will be extended to incorporate development of other modalities. For example, how motor development influences visual development and vise verse are the most interesting question. Adding other modalities such as tactile and auditory perceptions would alter interactions between infants and caregivers. Different responses from caregivers would cause changes in the developmental pathway of the MNS, which is another important issue to be addressed.

ACKNOWLEDGMENT

This study is partially supported by Grants-in-Aid for Scientific Research (Research Project Number: 22220002).

REFERENCES

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

[2] G. Rizzolatti *et al.*, *Mirrors in the brain: how our minds share actions, emotions*. Oxford University Press, USA, 2008.

[3] G. Pellegrino *et al.*, "Understanding motor events: a neurophysiological study," *Experimental brain research*, vol. 91, no. 1, pp. 176–180, 1992.

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

[5] M. Heiser *et al.*, "The essential role of Broca's area in imitation," *European Journal of Neuroscience*, vol. 17, no. 5, pp. 1123–1128, 2003.

[6] N. Nishitani and R. Hari, "Temporal dynamics of cortical representation for action," in *Proc. of the National Academy of Sciences of the United States of America*, vol. 97, no. 2, 2000, p. 913.

[7] 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.

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

[9] M. Brass and C. Heyes, "Imitation: is cognitive neuroscience solving the correspondence problem?" *Trends in Cognitive Sciences*, vol. 9, no. 10, pp. 489–495, 2005.

[10] S. Jones, "Infants learn to imitate by being imitated," in *Proc. of the 5th Intl. Conf. on Development and Learning*, 2006.

[11] Y. Kuniyoshi *et al.*, "From visuo-motor self learning to early imitation-a neural architecture for humanoid learning," in *Proc. of the 2003 IEEE Intl. Conf. on Robotics and Automation*, vol. 3, 2003, pp. 3132–3139.

[12] T. Chaminade *et al.*, "From self-observation to imitation: Visuomotor association on a robotic hand," *Brain research bulletin*, vol. 75, no. 6, pp. 775–784, 2008.

[13] M. Asada *et al.*, "Imitation learning based on visuo-somatic mapping," *Experimental Robotics IX*, pp. 269–278, 2006.

[14] T. Minato *et al.*, "A model of the emergence of early imitation development based on predictability preference," in *Proc. of the 9th Intl. Conf. on Development and Learning*, 2010, pp. 19–25.

[15] S. M. Mahler, *et al.*, *The Psychological Birth Of The Human Infant: Symbiosis And Individuation*. Basic Books, 7 1975.

[16] Y. Nagai, "The role of motion information in learning human-robot joint attention," in *Proc. of the 2005 IEEE Intl. Conf. on Robotics and Automation*, 2005, pp. 2069–2074.

[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] T. Minato *et al.*, "A baby robot platform for cognitive developmental robotics," in *Proceeding of Workshop on Synergistic Intelligence at the 2009 IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, 2009.

[19] J. Atkinson *et al.*, "Acuity and contrast sensivity of infant vision," *Nature*, vol. 247, pp. 403–404, 1974.

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

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

[22] 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.

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

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

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

[26] 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.

[27] L. B. Smith *et al.*, "Not your mother's view: the dynamics of toddler visual experience." *Developmental Science*, vol. 14, no. 1, pp. 9–17, 2011.