

Towards Robot Self-consciousness (I): Brain-Inspired Robot Mirror Neuron System Model and Its Application in Mirror Self-recognition

Yi Zeng^{1,2(✉)}, Yuxuan Zhao¹, and Jun Bai¹

¹ Institute of Automation, Chinese Academy of Sciences, Beijing, China
yi.zeng@ia.ac.cn

² Center for Excellence in Brain Science and Intelligence Technology,
Chinese Academy of Sciences, Shanghai, China

Abstract. Mirror Self-Recognition is a well accepted test to identify whether an animal is with self-consciousness. Mirror neuron system is believed to be one of the most important biological foundation for Mirror Self-Recognition. Inspired by the biological mirror neuron system of the mammalian brain, we propose a Brain-inspired Robot Mirror Neuron System Model (Robot-MNS-Model) and we apply it to humanoid robots for mirror self-recognition. This model evaluates the similarity between the actual movements of robots and their visual perceptions. The association for self-recognition is supported by STDP learning which connects the correlated visual perception and motor control. The model is evaluated on self-recognition mirror test for 3 humanoid robots. Each robot has to decide which one is itself after a series of random movements facing a mirror. The results show that with the proposed model, multiple robots can pass the self-recognition mirror test at the same time, which is a step forward towards robot self-consciousness.

Keywords: Robot self-consciousness · Mirror self-recognition · Mirror neuron system · Associative learning

1 Introduction

Self consciousness is of vital importance for an agent with real intelligence. Machine consciousness is a grand challenge for Artificial Intelligence research. In order to identify whether an animal species is with self consciousness, the Mirror Self-Recognition test is proposed [1]. Only a few animal species are considered to be with self consciousness. Besides human, animals that are considered to be with self-consciousness include: chimpanzees [1], orangutans [2], bonobos [3], gorillas [4, 5], Asiatic elephant [6], dolphins [7], orcas [8], Eurasian [9], etc.

Recent findings proofed the possibility of training rhesus monkeys to be with self consciousness [10]. With this possibility as a support, we hypothesize that

machine with a brain-inspired computational model can be trained to have self consciousness.

For mammalian brain, especially human brain, multiple brain regions are involved in self consciousness. They closely interact with each other and collectively form a comprehensive neural pathway. Two sub systems need to be paid more attention to. Namely, the Mirror Neuron System (MNS) [11], and the Cortical Midline Structures (CMS) [12]. The medial prefrontal cortex (MPFC) is a very important region in CMS for self consciousness [13]. The ventromedial prefrontal cortex (VMPFC) mostly responds to self, while the dorsomedial prefrontal cortex (DMPFC) primarily responds to others [14]. For the mirror self-recognition test, especially for a robot self-consciousness model, we hypothesize that mirror neuron system (MNS) learns the correlation of the original agent and the agent in the mirror, then MPFC is activated by the mirror neuron system to realize the robot self. Hence, creating a computational model for mirror neuron system (MNS) is a first preparation for realizing robot self-consciousness.

In order to create a robot with self-consciousness, in this paper, we propose a brain-inspired Robot Mirror Neuron System Model (Robot-MNS-Model) and we apply it to humanoid robots for mirror self-recognition. Although this investigation will be a long term exploration¹, the efforts in this paper try to have a step forward towards robot self-consciousness.

2 Brain-Inspired Robotic Mirror Neuron System Model

As a core architecture for the computational model of robot self-consciousness, we propose a Robotic Mirror Neuron System model (Robot MNS Model).

2.1 The Architecture of the Robotic Mirror Neuron System

The architecture of the Robot MNS Model is shown in Fig. 1. The model is mainly based on the understanding of human mirror neuron system introduced in [11, 15], and is with reconsideration to adapt to robotics.

The motion detection module receives visual inputs and detects motions in the visual sequence. It is composed of Extrastriate Body Area (EBA) and MT/V5. EBA is sensitive to human body and its parts, no matter they are static or moving [16]. MT/V5 is with the ability of motion detection, and responds to stimuli which are moving towards a certain direction with a certain speed [17]. In this model, EBA is with the function of body part detection, while MT/V5 is for orientation and speed detection of moving objects. Both EBA and MT/V5 transmit information to posterior superior temporal sulcus (pSTS). pSTS is sensitive to biological motion, and its function is to visually encodes biological motion [18]. The inferior parietal lobule (IPL) integrates visual inputs from pSTS and motion inputs from vPMC. Namely, IPL, which is one of the most important core area in mirror neuron system and for this model, does consistency checking on observed

¹ Robot Self-Consciousness Project: <http://bii.ia.ac.cn/robot-self>.

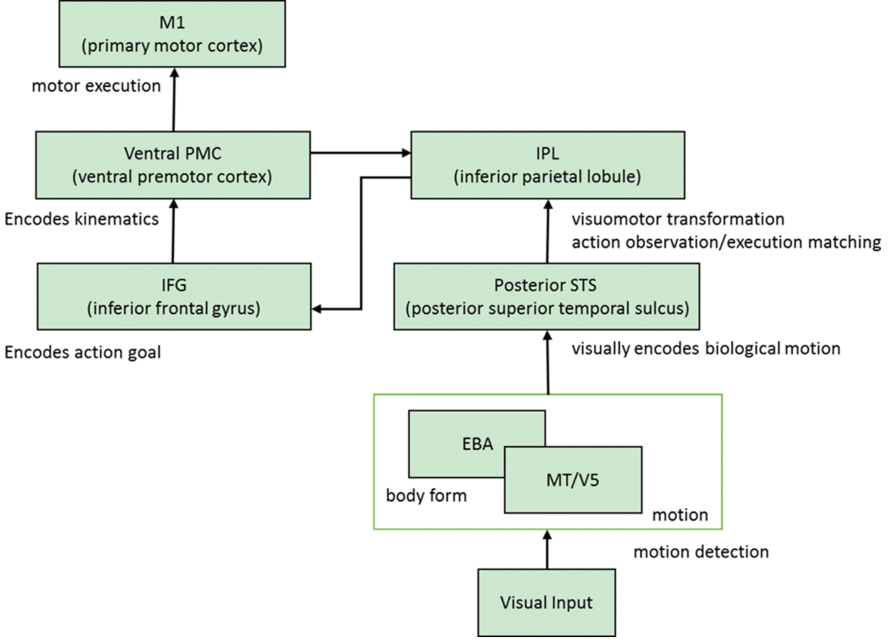


Fig. 1. The architecture of the robotic mirror neuron system model

action and motion execution [18,19]. The inferior frontal gyrus (IFG) encodes action goals, and it responds to goal driven motions [19]. In this model, IFG generates motion goals, and transmit information on motion goals to vPMC. It also make inference on motion intention based on inputs from IPL. The Ventral premotor cortex (Ventral PMC) encodes kinematics based on motion goal from IPL, the encoded information are sent to M1 for concrete motor execution, and to IPL for information integration. M1 encodes the strength and orientation of motion and controls the concrete motion execution [20].

In the robotic mirror neuron system model, motion goal generation and motion understanding are associated with IFG. The neural pathway of the proposed model is mainly based on understandings from [11,15,21,22]. In this model, there are mainly two pathways. Namely, The somato motion perception pathway (IFG \rightarrow vPMC \rightarrow IPL), and the visual motion perception pathway (visual inputs \rightarrow EBA & MT/V5 \rightarrow pSTS \rightarrow IPL). Both pSTS and vPMC send signals to IPL. If the signals from these two regions co-occur with each other in the same time slot (in this investigation, the time slot is consistent with the slot for STDP) and the sequences of movements are consistent with each other, new connections will be formed or existing connections will be strengthened to represent their consistency, then IPL activates mPFC and the robot recognizes the moving agent is itself. Figure 2 presents the associative learning process between pSTS and vPMC signals in IPL.

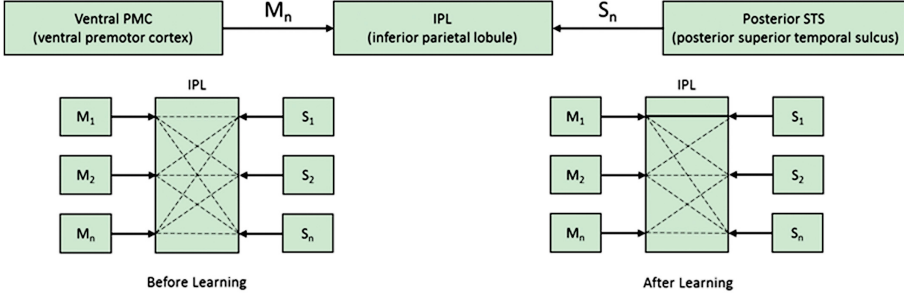


Fig. 2. Associative learning of pSTS and vPMC signals in IPL

2.2 Motion Execution and Somato Motor Perception

The inferior frontal gyrus (IFG) sends motion goals (such as moving left hand to a specific position) to ventral premotor cortex (vPMC). vPMC encodes the motion sequence and send the encoded information to M1, which is directly related to robot motion execution. At the same time, vPMC sends the motion sequence to IPL for consistency checking with visual motion perception.

2.3 Motion Detection

Here we propose a multiple brain region coordinated computational model for the visual dorsal pathway to simulate the cognitive function of motion detection. The model is a spiking neural network model, which is enlightened by the work introduced in [23].

As a preprocessing step, the image is firstly transformed to three spatial-temporal scales through convolutions with different Gaussian kernels. This is to simulate the effect of retina and can help to detect targets in different scales in the scene.

After preprocessing, the image is fed as input to V1. As is pointed out by neural scientists, most neurons in V1 related to motion can exhibit a crucial characteristic: direction selectivity. To clarify this, it is necessary to introduce the concept of spatial-temporal space. As a 2-D image, a resolution cell(or a pixel) can be located by its (x, y) spatial coordinates in the scene. However, for an image sequence, a third dimension t should be appended to describe the temporal change of this resolution cell. Thus, the image sequence can be described as a 3-D subspace with x , y , and t axis. A basic unit in motion can be represented as its spatial change of (x, y) with the temporal sequence t . In other words, this unit in motion is actually a trajectory in this 3-D subspace.

For a neuron related to motion in V1, direction selectivity means this neuron can reach its maximal response in a particular direction in the spatial-temporal space within its receptive field. This implies that this neuron responds to motion not simply in a spatial direction, but also in a specific speed. To simulate this physiological effect in a mathematical form, we sample the 3-D spatial-temporal

space in 28 dimensions, corresponding to 28 different motion directions and speeds. In fact, these directions correspond to different functional columns in V1.

As a further step, neurons from V1 are connected to MT. Different from the neurons in V1, the neurons in MT respond to motion in a specific direction, regardless of the speeds. Hence, this is a projection process from 3-D spatial-temporal space to 2-D spatial space. the input of MT neurons is in fact the linear transformation to the output of V1 neurons. In our simulation, we sample 8 spatial directions in MT, corresponding to motions in these 8 directions.

LIP takes the role of decision making in motion detection. Its function is similar to the output layer of artificial neural network. In [23], the output is the firing rates in 8 directions. In our investigation, however, what we concern is the degree of motion, regardless of the directions. To this end, we consider only the maximal firing rates of all the 8 directions. Since this maximal firing rate implies the intention of motion, we can use it to detect the targets in motion.

One key issue in the network is direction selectivity in V1. In [23], an algebra model is proposed, and the process can be briefly described as follows.

After preprocessing, the motion descriptors are calculated based on the input image sequence.

$$L_{kr}(x, y, t) = \alpha_{v1lin} \sum_{T=0}^3 \left[\sum_{Y=0}^{3-T} \left[\frac{3!}{X!Y!T!} (\hat{u}_{k,x})^X (\hat{u}_{k,y})^Y (\hat{u}_{k,t})^T \frac{\partial^3 f_r(x, y, t)}{\partial x^X \partial y^Y \partial t^T} \right] \right] \quad (1)$$

In this equation, X , Y and T are two spatial directions and the temporal dimension respectively. r is the scalar, which is 0, 1 or 2. k is one of the 28 spatial-temporal directions. α_{v1lin} is a tuning parameter. $L_{kr}(x, y, t)$ is the motion descriptor.

Then, the descriptors are translated to the responses of simple V1 cells.

$$S_{kr}(x, y, t) = \frac{\alpha_{filt \rightarrow rate, r} \alpha_{v1rect} L_{kr}(x, y, t)^2}{\alpha_{v1norm} \exp\left(\frac{-(x^2 + y^2)}{2\sigma_{v1norm}^2}\right) * \left(\frac{1}{28} \sum_{k=1}^{28} L_{kr}(x, y, t)^2\right) + \alpha_{v1semi}^2} \quad (2)$$

Here, σ_{v1norm} is the Gaussian kernel. $\alpha_{filt \rightarrow rate, r}$, α_{v1rect} , α_{v1norm} are parameters to be tuned. The responses of V1 simple cells are then further filtered as the responses of V1 complex cells.

$$C_{kr}(x, y, t) = \alpha_{v1comp} \exp\left(\frac{-(x^2 + y^2)}{2\sigma_{v1comp}^2}\right) * S_{kr}(x, y, t) \quad (3)$$

σ_{v1comp} is the Gaussian kernel. α_{v1comp} is the tuning parameter. This responses can be taken as the firing rate.

Enlightened by [24], we propose to use the classical Integrate and Fire (IF) model to generate spikes.

$$\begin{cases} \frac{dV}{dt} = G_{\theta}^{exc}(x_0, y_0, t)(E^{exc} - V(t)) \\ \quad + G_{\theta}^{inh}(x_0, y_0, t)(E^{inh} - V(t)) - g^L V(t) \\ \text{Spikes when } V = 1 \text{ and resets } V \text{ to } 0 \end{cases} \quad (4)$$

For the purpose of direction selectivity, the key is that the excitatory and inhibitory conductances are functions of the direction θ and spatial-temporal location (x, y, t) . For excitatory conductance:

$$G_{\theta}^{exc}(x, y, t) = (F_{\theta}^e + F_{\theta}^o) * L(x, y, t) \quad (5)$$

$L(x, y, t)$ is the input sequence. $*$ is the convolution operator. F_{θ}^e and F_{θ}^o can be describes as:

$$\begin{cases} F_{\theta}^e(x, y, t) = G_{\theta}^e(x, y)P_i(t) \\ F_{\theta}^o(x, y, t) = G_{\theta}^o(x, y)P_j(t) \end{cases} \quad (6)$$

G_{θ}^e and G_{θ}^o are two Gaussian kernels respectively. $P_i(t)$, $P_j(t)$ are subtractions of two Γ functions.

$$P_{\alpha}(t) = T_{\alpha, \tau}(t) - T_{\alpha+2, \tau}(t) \quad (7)$$

Here $T_{\alpha, \tau}(t)$ is the Γ function. The inhibitory conductance is described as follows.

$$G_j^{inh} = G_{max}^{inh} e^{-\frac{d_j^4}{2R^2}} \quad (8)$$

where j is the index of the neuron's spatial temporal neighbor. G_{max}^{inh} is a parameter, describing a maximal contribution of its neighbors. d_j is the spatial-temporal distance, and R is the radius of the scope that can contribute to the inhibitions.

3 Sensory-Motor Associative Learning

In the mirror neuron system, IPL integrates visual and motor information through associative learning. If motor information from vPMC match the visual information from pSTS, and the information are sent to IPL at the same time slot, then they are associated together in IPL. Associations are formed through Spike-Timing-Dependent-Plasticity (STDP) [25, 26]. The weight of association among visual inputs and motor outputs ($W_{\text{motor-visual}}$) is based on a set of calculations in Eq. 9.

$$\begin{aligned} \Delta W &= \begin{cases} A_+ \times e^{(\Delta t / \tau_+)} & \text{if } \Delta t < 0 \\ A_- \times e^{(\Delta t / \tau_-)} & \text{if } \Delta t \geq 0 \end{cases} \\ \Delta t &= t_{\text{motor}} - t_{\text{visual}} \\ W(t)_{\text{motor-visual}} &= W(t-1)_{\text{motor-visual}} + \Delta w_{\text{motor-visual}} \end{aligned} \quad (9)$$

ΔW is the adjustment function for STDP. A_+ and A_- are the maximum and minimum value of synaptic changes respectively. τ_+ and τ_- are time constants for synaptic updates. Δt is the time slot between the time for motor output (t_{motor}) and the time for visual recognition of movement (t_{visual}). In order to keep the biological plausibility, according to [25], $A_+ = 0.777$, $A_- = -0.237$, $\tau_+ = 16.8$ ms, $\tau_- = -33.7$ ms. If motor signals are transmitted to IPL before visual signals, then the synaptic connectivity will be strengthened, if visual signals come first, it will be weakened.

When there are not only one robot moving in front of a mirror, the robots need to decide which one is itself. The self recognition weight, denoted as $\text{self}_{\text{weight}}$, is proposed to evaluate which one is the specific robot itself.

$$\begin{aligned} \theta_{\text{predict}} &= \max W_{\text{motor}} \\ \text{Confidence}_i &= \begin{cases} 1 & |\theta_{\text{predict}} - \theta_{\text{visual}}| \leq \theta_{\text{threshold}} \text{ and } t_{\text{motor}} - t_{\text{visual}} < t_{\text{threshold}} \\ 0 & \text{otherwise} \end{cases} \\ \text{Self}_{\text{weight}} &= \left(\sum_{i=1}^n \text{Confidence}_i \right) / n \end{aligned} \quad (10)$$

θ_{predict} is the predicted angle based on motor information. If θ_{predict} and θ_{visual} are close to each other within a certain threshold ($\theta_{\text{threshold}}$), at the same time, t_{motor} is before t_{visual} and they are close to each other within the time slot $t_{\text{threshold}}$, the confidence value for the robot under state i is (confidence_i) is 1. $\text{Self}_{\text{weight}}$ is the average value for all the n states during robot movements.

4 Robots Mirror Self-recognition Test

In order to validate the proposed Brain-inspired Robot Mirror Neuron System model, we deploy the computational model to humanoid robotics and challenge the model with the robots mirror self-recognition test. Three robots are required to recognize itself and distinguish itself from others in front of a mirror. The prior knowledge for robots are as the following: If two robots are on its right, then it is in Position 1. If two robots are on its left and right respectively, then it is Position 2. If two robots are on its left, then it is in Position 3. Hence, the judgement process can be described as Eq. 11. During the mirror test, they are required to identify which position it belongs to, and in this way, it obviously needs to know which one is itself first.

$$\text{Position ID} = \begin{cases} 1 & \text{LeftCount}=0 \ \& \ \text{RightCount}=2 \\ 2 & \text{LeftCount}=1 \ \& \ \text{RightCount}=1 \\ 3 & \text{LeftCount}=2 \ \& \ \text{RightCount}=0 \end{cases} \quad (11)$$

The robots are assigned random movements if they do not have any obvious solution to a specific task. All of their sensory functions (vision, audition) and



Fig. 3. Visual inputs and motion detection for robots mirror self-recognition

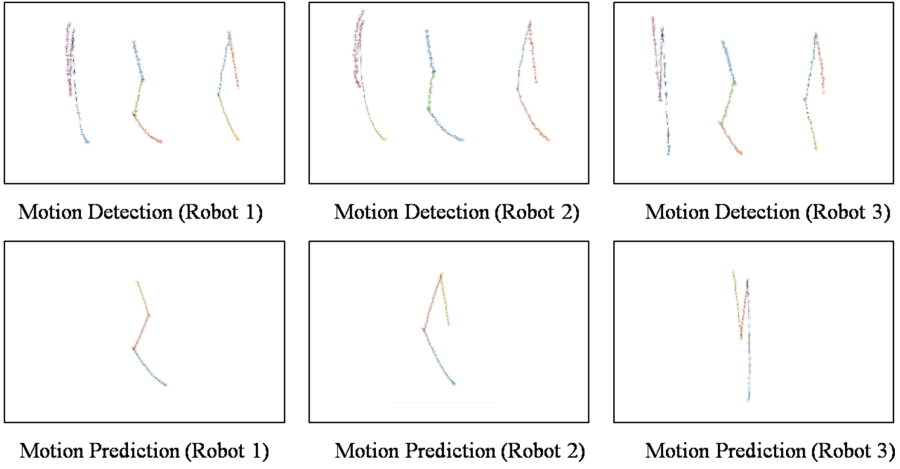


Fig. 4. Motion detection and motion prediction by different robots

motor function are activated with attempts to achieve the goal. Since the robots are in front of a mirror, and their vision systems are active, functions of their visual systems are active. Hence, motion detection from dorsal ventral pathway is functioning. Figure 3 presents the visual inputs of robots and their motion detection.

Robots will keep random movement until they can confirm their positions (i.e. identify which one it is). Figure 4 presents a sample for motion detection of their moving hands through visual inputs of the three robots and motion pathway prediction based on each robot's actual motor outputs.

Figure 5 presents the beliefs of each robots after each random movement. The darkness of each square is negative relevant to the belief values. After movement 2, Robot 2 is not aware which one is itself, while after movement 3, each robot can identify themselves in the mirror.

In order to test the proposed robot mirror neuron system model, 9 sets of robot mirror self-recognition tests are made. For each test, the position of the three robots are changed. Hence they need to re-recognize themselves after each

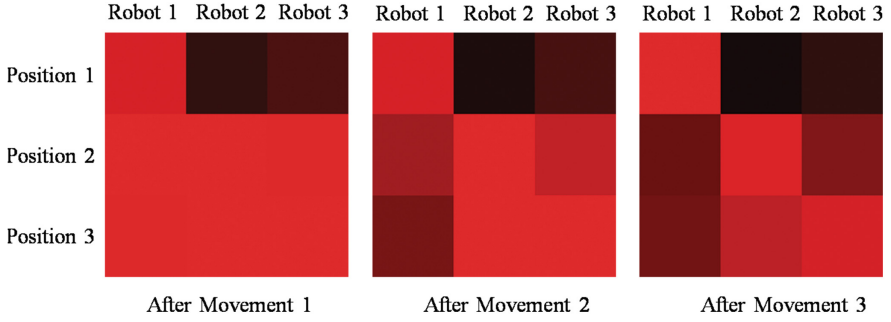


Fig. 5. Beliefs of their positions after different random movements

test. The three robots successfully passed all the 9 tests. Videos on the robots mirror self-recognition test is available at the Robot Self-consciousness Project page².

5 Conclusion

With the long term goal of building a robot with consciousness, especially with self-consciousness, as a first step, this paper provides an attempt to build a brain-inspired robot mirror neuron system model. Then we apply the model to robots with the purpose of passing the mirror self-recognition test. The evaluation indicates that the proposed model is biologically plausible and computationally feasible as a core component for robot self consciousness. Mirror Neuron System is only part of the neural pathway for self consciousness.

In order to provide a more comprehensive model and increase the level of self consciousness for robots, our current and future work is to extend the model to involve more regions and pathways that are relevant to self consciousness.

Acknowledgment. This study was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB02060007), and Beijing Municipal Commission of Science and Technology (Z151100000915070, Z161100000216124).

References

1. Gallup, G.G.J.: Chimpanzees: self recognition. *Science* **167**(3914), 86–87 (1970)
2. Suarez, S.D., Gallup, G.G.J.: Self-recognition in chimpanzees and orangutans, but not gorillas. *J. Hum. Evol.* **10**(2), 175–188 (1981)
3. Walraven, V., van Elsacker, L., Verheyen, R.: Reactions of a group of pygmy chimpanzees (*Pan paniscus*) to their mirror-images: evidence of self-recognition. *Primates* **36**(1), 145–150 (1995)

² Robot Self-Consciousness Project: <http://bii.ia.ac.cn/robot-self>.

4. Patterson, F.G.P., Cohn, R.H.: Self-recognition and self-awareness in lowland gorillas. In: *Self-Awareness in Animals and Humans: Developmental Perspectives*, pp. 273–290. Cambridge University Press (1994)
5. Posada, S., Colell, M.: Another gorilla recognizes himself in a mirror. *Am. J. Primatol.* **69**(5), 576–583 (2007)
6. Plotnik, J.M., Waal, F.D., Reiss, D.: Self-recognition in an Asian elephant. *Proc. Natl. Acad. Sci.* **103**(45), 17053–17057 (2006)
7. Marten, K., Psarakos, S.: Evidence of self-awareness in the bottlenose dolphin (*Tursiops truncatus*). In: *Self-Awareness in Animals and Humans: Developmental Perspectives*, pp. 361–379. Cambridge University Press (1994)
8. Delfour, F., Martenb, K.: Mirror image processing in three marine mammal species: killer whales (*Orcinus orca*), false killer whales (*Pseudorca crassidens*) and California sea lions (*Zalophus californianus*). *Behav. Process.* **53**(3), 181–190 (2001)
9. Prior, H., Schwarz, A., Gntkrn, O.: Mirror-induced behavior in the magpie (*Pica pica*): evidence of self-recognition. *PLOS Biol.* **6**(8), e202 (2008)
10. Chang, L., Fang, Q., Zhang, S., Poo, M., Gong, N.: Mirror-induced self-directed behaviors in rhesus monkeys after visual-somatosensory training. *Curr. Biol.* **25**(2), 212–217 (2015)
11. Iacoboni, M., Dapretto, M.: The mirror neuron system and the consequences of its dysfunction. *Nat. Rev. Neurosci.* **7**(12), 942–951 (2006)
12. Northoff, G., Heinzel, A., de Greck, M., Bermpoh, F., Dobrowolny, H., Panksepp, J.: Self-referential processing in our brains: a meta-analysis of imaging studies on the self. *NeuroImage* **31**, 440–457 (2006)
13. Heatherton, T.F.: Neuroscience of self and selfregulation. *Ann. Rev. Psychol.* **62**, 363–390 (2011)
14. Denny, B.T., Kober, H., Wager, T.D., Ochsner, K.N.: A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J. Cogn. Neurosci.* **24**(8), 1742–1752 (2012)
15. Thakkar, K.N., Peterman, J.S., Park, S.: Altered brain activation during action imitation and observation in schizophrenia: a translational approach to investigating social dysfunction in schizophrenia. *Am. J. Psychiatry* **171**(5), 539–548 (2014)
16. Peelen, M.V., Wiggett, A.J., Downing, P.E.: Patterns of fmri activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron* **49**(6), 815–822 (2006)
17. Perrone, J.A., Thiele, A.: Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nat. Neurosci.* **4**(5), 526–532 (2001)
18. Grossman, E.D., Blake, R.: Brain areas active during visual perception of biological motion. *Neuron* **35**(6), 1167–1175 (2002)
19. Hamzei, F., Vry, M.S., Saur, D., Glauche, V., Hoeren, M., Mader, I., Weiller, C., Rijntjes, M.: The dual-loop model and the human mirror neuron system: an exploratory combined fMRI and DTI study of the inferior frontal gyrus. *Cereb. Cortex* **26**(5), 2215–2224 (2016)
20. Georgopoulos, A.P., Schwartz, A.B., Kettner, R.E.: Neuronal population coding of movement direction. *Science* **233**(4771), 1416–1419 (1986)
21. Sasaki, A.T., Kochiyama, T., Sugiura, M., Tanabe, H.C., Sadato, N.: Neural networks for action representation: a functional magnetic-resonance imaging and dynamic causal modeling study. *Front. Hum. Neurosci.* **6**, 236 (2012)
22. Mehta, U.M., Thirthalli, J., Aneelraj, D., Jadhav, P., Gangadhar, B.N., Keshavan, M.S.: Mirror neuron dysfunction in schizophrenia and its functional implications: a systematic review. *Schizophrenia Res.* **160**(1–3), 9–19 (2014)

23. Beyeler, M., Richert, M., Dutt, N.D., Krichmar, J.L.: Efficient spiking neural network model of pattern motion selectivity in visual cortex. *Neuroinformatics* **12**(3), 435–454 (2014)
24. Escobar, M.J., Wohrer, A., Kornprobst, P., Vieville, T.: Biological motion recognition using a MT-like model. In: *Proceedings of the 3rd IEEE Latin American Robotic Symposium*, pp. 47–52 (2006)
25. Bi, G., Poo, M.: Synaptic modification by correlated activity: Hebb’s postulate revisited. *Annu. Rev. Neurosci.* **24**, 139–166 (2001)
26. Song, S., Miller, K.D., Abbott, L.F.: Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* **3**(9), 919–926 (2000)

Advances in Brain Inspired Cognitive Systems
8th International Conference, BICS 2016, Beijing, China,
November 28-30, 2016, Proceedings
Liu, C.-L.; Hussain, A.; Luo, B.; Tan, K.C.; Zeng, Y.;
Zhang, Z. (Eds.)
2016, XIII, 368 p. 159 illus., Softcover
ISBN: 978-3-319-49684-9