Prediction Error in the PMd As a Criterion for Biological Motion Discrimination: A Computational Account

Yuji Kawai, Yukie Nagai, *Member, IEEE*, and Minoru Asada, *Fellow, IEEE*

***Abstract*—Neuroscientific studies suggest that the dorsal premotor area is activated by biological motions, and is also related to the prediction errors of observed and self-induced motions. We hypothesize that biological and nonbiological motions can be discriminated by such prediction errors. We therefore propose a model to verify this hypothesis. A neural network model is constructed that learns to predict the velocity of the self's next body movement from that of the present one and produces a smooth movement. Consequently, a property of the input sequence is represented. The trained network evaluates observed motions based on the prediction errors. If these errors are small, the movements share a representation with the self-motor property, and therefore, are regarded as biological ones. To verify our hypothesis, we examined how the network represents the biological motions. The results show that predictive learning, supported by a recurrent structure, helps to obtain the representation that discriminates between biological and nonbiological motions. Moreover, this recurrent neural network can discriminate the ankle and wrist trajectories of a walking human as biological motion, regardless of the subject's sex, or emotional state.**

***Index Terms*—Biological motion, dorsal premotor area (PMd), one-third power law, predictive learning, recurrent neural network.**

I. INTRODUCTION

ONE OF the amazing abilities of young infants is that they can discriminate between biological and nonbiological motions. Many previous studies have shown that biological motion perception is processed in the occipital and temporal brain regions (e.g., [1]–[4]). In contrast, it has been reported that the frontal and parietal cortices also activate during the observation of biological motions (e.g., [5] and [6]). However, the role and mechanism of the activation of the frontal and parietal areas are not well understood. In this paper, we propose a computational model for biological motion discrimination in the motor-related area suggested by neuroscientific, developmental, and psychological findings.

We focus on the fact that the dorsal premotor area (PMd) activates when a biologically moving object is observed [6]–[8]. Casile *et al*. [8] used the one-third power law to define biological motion, where the tangential speed is proportional to the one-third power of the radius of curvature (i.e., the speed decreases with larger curvature) [9]. A recent electroencephalographic study also reported that brain oscillation in the motor-related area is modulated by the observation of movements following the law [10]. This law can be observed in many biological motions, e.g., hand movements [9], eye movements [11], and ankle movements during walking [12], [13]. Thus, it is known as the kinematic invariance of biological motions. Even mentally simulated motor trajectories, which are assumed to originate in motor planning within motor-related areas, comply with the law [14]. It has been shown that such smooth movements are recognized as biologically natural human movements [15], [16]. Therefore, we suppose that the PMd may represent the one-third power law, which enables infants to find biological motions.

Interestingly, this ability may not require visual learning. Neonates have few visual experiences, but are able to detect biological motions [17], [18]. These studies proposed that this ability may be innate. However, we suppose that it is still possible to acquire this ability through motor experience during the fetal period. It was observed that fetuses produce smooth and voluntary reaching actions [19] and neonates' hand movements obey a one-third power-like law [20]. This implies that neonates obtain the one-third power law through motor experiences before their visual experiences. The obtained law is then applied to visual information process in order to perceive biological motions.

The PMd generally plays an important role in producing reaching actions: it activates during planning, execution, imagination, and observation of reaching actions [21], [22]. The following lists are several findings about these functions.

1) An anatomical connection with the superior parietal lobule (SPL) [23], and an interaction between the PMd and SPL enables people to accurately perform reaching actions [24].
2) The reception of the spatial information of a target object and of an actor's body from the SPL enables them to produce a desired trajectory [24].

Manuscript received May 23, 2016; revised November 28, 2016; accepted February 7, 2017. Date of publication February 13, 2017; date of current version June 8, 2018. This work was supported in part by the JSPS Grants-in-Aid for Specially Promoted Research under Grant 24000012, and in part by the Grants-in-Aid for JSPS Fellows under Grant 13J00756.

The authors are with the Graduate School of Engineering, Osaka University, Osaka 565-0871, Japan (e-mail: kawai@ams.eng.osaka-u.ac.jp; yukie@ams.eng.osaka-u.ac.jp; asada@ams.eng.osaka-u.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCDS.2017.2668446

3) The PMd reacts to errors in the predicted trajectories of one's own limb [25] and those of others' reaching motions [26], and activates during the prediction of an object's motion and even its geometric shape, as well [27]–[29].

These findings may suggest that prediction errors in self-induced motions, others' motions, and object's motions can discriminate between biological and nonbiological motions, even when the observed target is not an agent.

The above suggestion leads to the following hypothesis: a neural network in the PMd represents biological motion, i.e., the one-third power law acquired from one's own motor experiences without any visual information. Based on the acquired representation, this network predicts others' observed movements. It then regards them as biological motions if their prediction errors are small, i.e., they share the representation with self-induced movements. To verify this hypothesis and clarify the requirements for a neural network to represent the biological law, we constructed a computational model. We utilize the velocity of an arm movement as an input of the network because a monkey's neurons in the PMd represent the velocities of self-induced arm movements (e.g., [30]). The one-third power law includes curvature, which is a function of acceleration. Therefore, the network must store past input and implicitly represent acceleration information to obtain the law. Predictive learning may encourage the organization of a neural network toward such information processing.

In the first experiment, we examine whether learning motor prediction allows the network to represent the law. Next, we input point light displays (PLDs) of human walkers to the network trained using the self-induced movements and show that the model can discriminate them from nonbiological motions. This task corresponds to the experiments for neonates (e.g., [17]). Finally, we attempt to simulate the experiment by Beets *et al*. [31] who showed that motor training without vision improved the visual motion discrepancy (details are given in the next section). Similar to the training in [31], our model is trained using three kinds of motions that either do or do not follow the one-third power law, and then its discrimination ability is evaluated. The measure is how clearly the hidden neurons are separated according to the input discrepancy.

This paper is partially based on our previous paper [32]. The basic idea is almost the same, and some experiment results overlap. However, in this paper, we analyze the internal representation of the learned neural networks in experiment 1 and show that our model is able to explain the experiment by Beets *et al*. [31] in experiment 4.

II. RELATED STUDY

There are several modeling and psychological studies related to biological motion representation. They are based on visual information processing [33], learning sensorimotor prediction [34]–[37], an RNN model limited to reaching motion [38], and a conceptual model [39], [40]. Our model outperforms all these models, and further behavioral evidence [31] supports our model.

Giese and Poggio [33] proposed a computational model for the recognition of PLDs. This model explains the information processing in the visual cortex from the primary visual area to the superior temporal sulcus (STS). Along the visual path, local motion information, observed through a receptive field-like mechanism, is integrated into global motion information. The global information is eventually associated with a gait pattern label, e.g., walking or running, in a supervised manner. However, this model does not consider nonvisual motor learning as neonates may perform [17], [18], [41]. Conversely, in our model, the motor-related area detects biological motions through motor experiences alone based on motor prediction.

There are many computational and robotic studies on cognitive models employing sensorimotor prediction (e.g., [34]–[37]). Most of them first learn associations between visual and motor (or proprioceptive) information during self-induced movements. The learned associations are then applied to visual information during the observation of another person's movement and predict its motor information or future visual information. This mechanism enables the models to predict the actor's goal or motor intention [34], [35]. Furthermore, the model can discriminate between self-other movements based on the prediction: movements with small prediction errors are evaluated as self-induced movements [36]. Schrodt *et al*. [37] utilized the visuomotor associations for biological motion perception. Their model recognized observed actions as self-induced motor patterns and anticipated their future motions. These models require associative learning with vision, while our model does not need vision because it focuses on the universally observed kinematic law of human movements.

Sawaragi and Kudoh [38] reported that predictive learning for reaching actions enables an artificial neural network to represent the actions. A recurrent neural network (RNN) was fed with the time-series positions of a hand and arm during reaching actions and trained to predict the one-ahead positions from the current positions. Consequently, the hidden neurons in the RNN represented angles of the arm and horizontal positions of the hand. However, the position data were limited to a reaching context, and therefore the representations were specific only for reaching movements. Our aim is to represent more general biological motions, i.e., the one-third power law, regardless of the modalities (motor or vision). Further, our model complies with the neuroscientific evidence for motor prediction and error detection in the PMd, while the previous model did not consider them.

The key idea that neonates may have biological motion representation is similar to the active intermodal mapping (AIM) model proposed by Meltzoff and Moore [39], [40]. In addition to the above-mentioned neonates' ability to discriminate biological motions, it is reported that they can imitate others' facial expressions [39]. The AIM model implies that the neonatal imitation requires visuomotor representation for facial movements: motor (proprioceptive) information of their own face and visual information of another's face are integrated into the visuomotor representation, and the neonates obtain the correspondence between them by extracting their common representation. Although the AIM model is similar to our model in terms of use of the visuomotor representation, this model is conceptual, and therefore, has not been implemented as a computational learning process. In contrast, we propose a computational model and aim to propose a neural basis for a learning process to acquire this representation.

There is further evidence to support the biological motion representation learned through nonvisual motor experiences. Beets *et al.* [31] showed that motor training improves the performance of visual discrimination for learned movements. Blindfolded adult subjects were trained to make hand movements that violated the one-third power law. After the motor training, they answered questions about whether two consecutive visual stimuli were the same. The stimuli were movements of a dot following different velocity profiles that violated the one-third power law. Performance of the task was improved depending on the success of the motor training. This result suggests that the motor representation acquired through motor experiences can be applied to the visual discrimination ability. This visuomotor representation may be extended to movements that violate the biological law as well.

III. MODEL FOR BIOLOGICAL MOTION DISCRIMINATION

The structure of the proposed model is shown in Fig. 1. The red box encloses a neural network used to represent biological motion. This network corresponds to the PMd. The blue box corresponds to the SPL/visual areas for coding visual inputs. This model is first trained with desired trajectories to produce smooth hand movements. The motor input consists of time-series velocities, and the neural network learns to output the next input, shown as solid arrows in Fig. 1. This motor-based learning is expected to enable the network to represent self-induced biological motions, i.e., the one-third power law. After the motor experience, the neural network is given visual information from the SPL and estimates its prediction error, shown as broken arrows in Fig. 1. If the errors are small, i.e., if the observed motions share the representation with self-induced motions, the model regards them as biological motions; otherwise, it regards them as nonbiological ones.

It is assumed that the planning of human-like smooth trajectories is modeled by minimizing the jerk, i.e., the derivative of acceleration [42]. Note that the minimum-jerk trajectories obey the one-third power law [43], but the law does not always satisfy minimum jerk. Fig. 2 shows an example of the minimum-jerk trajectory. As we can see, it exhibits a biological velocity profile such that the speed is slower when the curvature is greater, i.e., around points B and C.

The input signals to the neural network are time-series velocities of the minimum-jerk trajectories. At time t, a velocity [direction $\theta(t)$ [rad], and magnitude $v(t)$] is transformed into the activations of eight neurons that have direction selectivity at intervals of $\pi/4$ radians. Thus, the population-coded velocity is given as

$$\boldsymbol{V}_m(t) = [v_1(t)v_2(t)\cdots v_8(t)]$$

$$v_i(t) = v(t)\exp\left(-\frac{\left(\theta(t)-\frac{\pi}{4}i\right)^2}{2\sigma^2}\right) \quad (i = 1,\ldots,8) \qquad (1)$$

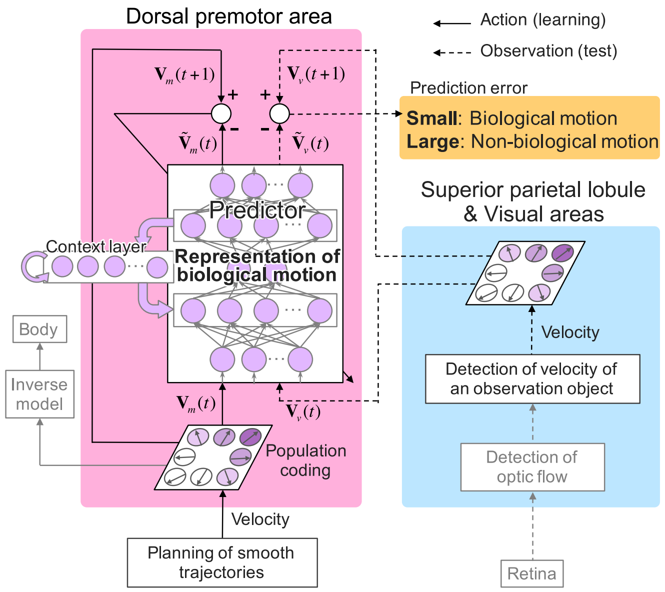
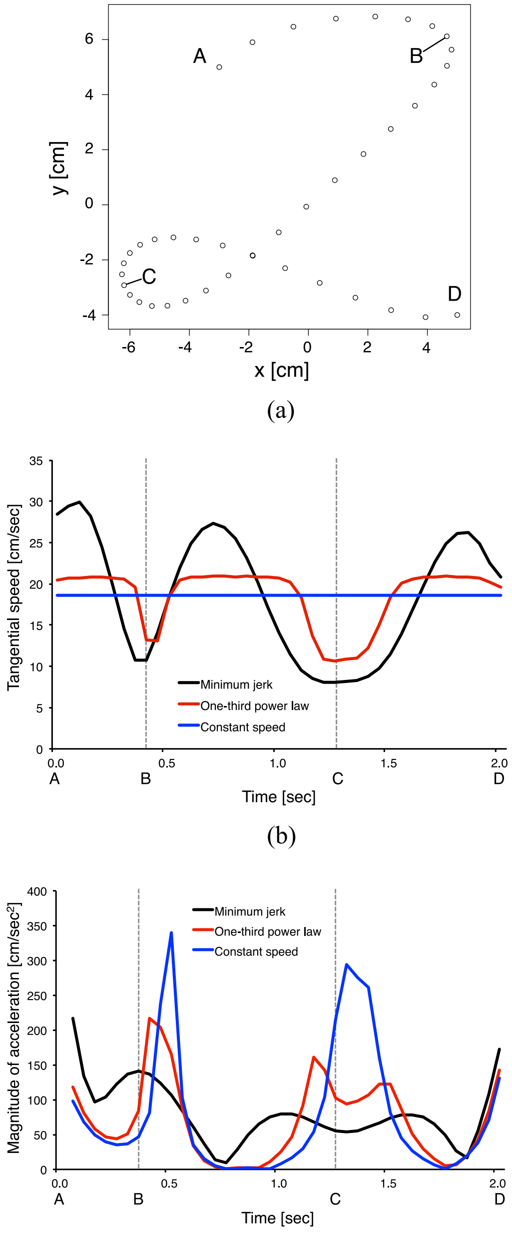


Fig. 1. Model for biological motion detection in the PMd. The neural network is trained to predict the velocity at the next time (solid arrows in the red box) and is expected to represent biological motion. Visual information is then fed into the learned network, which uses prediction error to evaluate how biological this information is (the broken arrows).

where σ is a constant. Velocity $\boldsymbol{V}_m(t)$ is fed into the neural network that learns to minimize the error between its output $\tilde{\boldsymbol{V}}_m(t)$ and subsequent input $\boldsymbol{V}_m(t+1)$. This predictive learning should allow the model to represent a property in the inputs.

After the predictive learning, the network evaluates whether the observed motions are biological or not. The time-series velocities of an observed object are coded in the same manner as motor information [i.e., as $\boldsymbol{V}_v(t)$], which are inputs of the network. The model estimates prediction errors by comparing output $\tilde{\boldsymbol{V}}_v(t)$ with subsequent input $\boldsymbol{V}_v(t+1)$, that is, $\text{Error} = (\boldsymbol{V}_v(t+1) - \tilde{\boldsymbol{V}}_v(t))^2$. If the observed motions share the representation of self-induced movements, the prediction errors are small. Therefore, the model judges the observed motions that have small errors to be biological motions.

The neural network that we used as a predictor is a five-layer sandglass-type network [44], in which the number of neurons in the middle layer is less than those in the outer layers (the center of the red box in Fig. 1). The lower dimensionality on the path to the output layer enables the neurons to acquire more compact representation of the input data [45], [46]. Further, this compact representation often avoids over-fitting and improves the network's generalization. We add a recurrent structure, an RNN, to the network to store the past input information. The neural activations of the fourth layer are returned to the second layer through the context layer. This additional layer allows the network to consider the past and current velocities to yield acceleration-like information. This may be useful for representing the one-third power law because it is a function of acceleration, and the effect of this structure is investigated in the experiments. The output function of the neurons in the first, third, and fifth layers is linear, i.e., $y = \lambda u$, while the function of the neurons in the second, fourth, and context layers is sigmoid, i.e., $y = [1/(1+\exp(-\lambda u))]$, where



(a)

(b)

(c)

Fig. 2. Example of trajectories given initial and terminal conditions. (a) Positions of a minimum-jerk trajectory in intervals of 0.05 s. This trajectory is generated from the initial condition [position $\boldsymbol{x} = (-3, 5)$ cm, velocity $\boldsymbol{v} = (18, 22)$ cm/s and acceleration $\boldsymbol{a} = (210, -170)$ cm/s^2 at time $t = 0.0$ s], the via condition [$\boldsymbol{x} = (-1, -1)$ cm and $\boldsymbol{v} = (-18, -18)$ cm/s at $t = 0.8$ s], and the terminal condition [$\boldsymbol{x} = (5, -4)$ cm, $\boldsymbol{v} = (20, 6)$ cm/s and $\boldsymbol{a} = (-60, 200)$ cm/s^2 at $t = 2.0$ s]. (b) Tangential speed. (c) Magnitude of acceleration of minimum-jerk (black), one-third power law (red), and constant tangential speed (blue) trajectories. The paths of these trajectories are the same, but their velocity and acceleration profiles are different. A–D in plots (b) and (c) corresponds to point A–D in plot (a).

y and u indicate the output and activity of a neuron, respectively. This activity is given as a summation of the products of the connection weights and output from the previous layer. The connection weights between the fourth and context layers and the self-recursive weights in the context layer remain constant throughout the learning, whereas the others are learned by standard back-propagation. This type of RNN was also employed by the above-mentioned previous model to represent reaching action [38]. In addition, many other models of the motor areas assume a recurrent structure to deal with time-series motor and sensor information (e.g., [37], [47], and [48]). Therefore, the RNN seems to be a reasonable model for learning and controlling motor behaviors in the motor area.

We employ the population-coded velocity as both the motor and visual inputs in accordance with neurophysiological studies. It is known that monkey's neurons with direction selectivity in the PMd code the velocities of hand movements [30], [49]. Further, a population coding method similar to ours (1) can decode a monkey's neural activities in the PMd during their hand movements. Consequently, smooth movements obeying the one-third power law can be successfully reconstructed using the population coding [50], [51]. Neurons in the middle temporal (MT) area represent the velocity of an observed object in a population-coding manner [52]. Therefore, population coding is a general method for representing movements, regardless of modality.

IV. EXPERIMENTAL PROCEDURE

A. Experiment 1: Acquisition of Motor Representation

In the first experiment, we investigated how a neural network represents biological motion through self-induced motor experiences. The neural network was first trained with five minimum-jerk trajectories [42] as desired movements. Next, ten unlearned minimum-jerk trajectories were fed into the trained network, which corresponds to the observation of biological motion. We additionally tested other types of trajectories (according to the one-third power law, but not minimum jerk[1] or constant tangential speed) for comparison. In fact, the output errors of the unknown minimum-jerk trajectories would be smaller than those of the constant-speed inputs if the network successfully learned the minimum-jerk criterion. Note that the errors of the nonminimum-jerk trajectories that obey the one-third power law were also expected to be smaller. That is, owing to the general property of neural networks (interpolation and extrapolation), the representation acquired from the minimum-jerk trajectories may be able to cope with the one-third power law trajectories, suggesting that it could be a biological motion representation.

To verify our model, we designed comparative experiments with two types of learning targets and two types of network structures (i.e., four experimental conditions). Regarding the learning targets, a neural network learned to minimize prediction errors [i.e., error $= (\tilde{\boldsymbol{V}}(t) - \boldsymbol{V}(t+1))^2$] or the output errors of auto-association (identity mappings) [i.e., error $= (\tilde{\boldsymbol{V}}(t) - \boldsymbol{V}(t))^2$]. This comparison suggests the importance of motor prediction in the PMd. Concerning the network structures, we compared the results of networks with and without a recurrent structure, i.e., an RNN and a feedforward network (FNN), respectively.

The inputs for the motor-related learning were limb movements given by a minimum-jerk model in the 2-D space [42]. The positions, velocities, and accelerations on the initial, via,

---

[1] Although the minimum jerk is a subset of the one-third power low, we refer to the one-third power low as it excluding the minimum jerk hereafter.

TABLE I
PARAMETERS OF THE RNN AND ITS LEARNING

Parameter	Value
Number of neurons (1st, 2nd, 3rd, 4th, 5th layers)	(8, 15, 2, 15, 8)
σ in Eq. (1)	60
λ in the output function	0.8
Fixed weights for the self-recursion in the context layer	0.8
Fixed weights from the forth layer to the context layer	1
Range of initial weights (random assignment)	$[-0.5, 0.5]$
Learning rate	0.01
Learning epochs	10,000

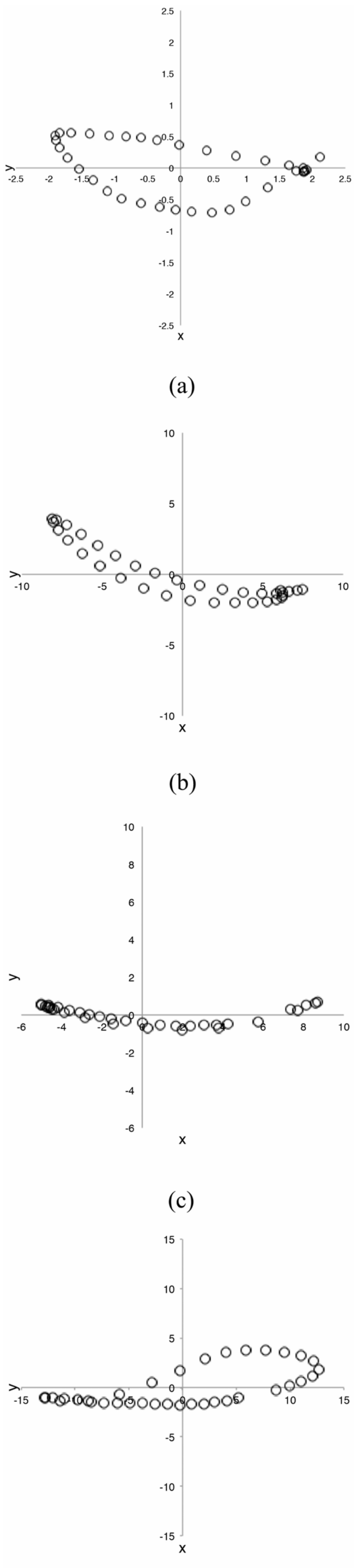
and terminal points were given in a random manner. All movement durations were 2.0 s, and the sampling rate was 20 Hz (0.05 s intervals). At every sampling time, the velocity of the trajectory was coded as a neural population (1). The 8-D vectors were normalized from 0 to 1 and were then fed into the network.

Unknown minimum-jerk trajectories were used to test the trained network. We generated them the same way as for the motor-based learning. The one-third power law and constant-speed trajectories were generated while the paths of the trajectories and duration time were the same as those of the minimum-jerk trajectories. Fig. 2(b) and (c), respectively, and magnitude of the acceleration of the minimum-jerk trajectory (the black line), one-third pow show the tangential speed er law trajectory (the red line), and constant-speed trajectory (the blue line). As we designed, the paths of their trajectories were exactly the same as the one shown in Fig. 2(a), regardless of the time stamp of each small circle. Fig. 2(b) shows that the minimum-jerk and one-third power law trajectories had a similar velocity profile that slowed at the curves (around B and C). However, the acceleration of the minimum-jerk trajectory is very smooth, while that of the one-third power law trajectory is not [see Fig. 2(c)]. Therefore, to represent biological motions from the minimum-jerk trajectories, the network should avoid over-fitting such smooth acceleration.

The parameters of the RNN are summarized in Table I. These parameter settings were used in all experiments. We evaluated each trajectory based on its average squared error.

B. Experiment 2: Learning From Acceleration and Velocity

In the above experiment, we employed the velocities of limb movements as the network input based on neurophysiological research (e.g., [30] and [49]). The velocity alone at one instant in time is insufficient to compute the curvature, which requires the acceleration. Recall that the one-third power law represents the relationship between the curvature and speed. Therefore, we supposed that predictive learning supported by a recurrent structure needs to internally represent the acceleration required for the one-third power law. However, explicitly giving the acceleration to the network seems to facilitate acquisition of the law even without predictive learning. Here, we gave the accelerations as well as the velocities of the minimum-jerk trajectories to the network in the learning and observation phases.



(a)

(b)

(c)

(d)

Fig. 3. Body part trajectories of a man walking in a neutral emotional state, reproduced from our previous study [32]. (a) Elbow. (b) Wrist. (c) Knee. (d) Ankle.

The accelerations were computed by subtracting the consecutive velocities

$$a_x = (v_x(t) - v_x(t-1))/0.05. \quad (2)$$

They were coded similarly to the velocities. The input vectors had 16-D (eight for the velocity and eight for the acceleration), and the number of the output neurons was also 16. The network learned prediction or auto-association of the inputs (both accelerations and velocities). The other conditions were the same as those of the first experiment.

C. Experiment 3: Recognition of Human Walking PLDs

We examined whether the network trained with the minimum-jerk trajectories in the first experiment could detect real biological motion. We used the PLDs of 28 walkers (14 men and 14 women aged from 17 to 28 years; the mean age is 22.5 years) from a database of 3-D PLDs [53]. Each walker has one of four emotional states: neutral, angry, happy, or sad. Each point on a body was normalized to fix the waist and was projected on a 2-D space as viewed from the side of the walker, which is like the presentation of a general stimulus in psychological experiments. We only used four points on the elbow, wrist, knee, and ankle because of their large displacements. The sampling frequency, normalization, and population coding method were the same as those in experiment 1. Fig. 3 shows an example trajectory of each body part of a male walker in a neutral emotional state.

D. Experiment 4: Discrimination of Input Patterns

In the final experiment, we aimed to confirm the result by Beets *et al.* [31]. This result indicated that the motor learning of hand movements that do not satisfy the one-third power law can improve the discrimination performance of observed movements that have the same successfully learned velocity patterns. We defined the discrimination performance to be the degree of classification of the neural activations in the compressive layer, i.e., the third layer. A network was trained using three kinds of trajectory patterns: minimum jerk, constant speed, and a minus one-third power law. The first pattern satisfies the one-third power law, but the others do not. Note that trajectories obeying the minus one-third power law exhibit a kinematic property opposite to those of the one-third power law: the speed increases with larger curvature. Five trajectories for each velocity pattern were used. Unknown trajectories with the same velocity patterns were then fed into the learned network. The activities of the neurons in the compressive layer were analyzed to determine if they could discriminate between the three velocity patterns.

We used a class separation index J to evaluate the degree of discrimination

$$J = \mathrm{tr}\left(\Sigma_W^{-1}\Sigma_B\right) \tag{3}$$

where $\mathrm{tr}(A)$ indicates the diagonal sum of a matrix A. Here, Σ_W and Σ_B are intraclass and interclass covariance matrices, respectively, given as

$$\Sigma_W = \frac{1}{n}\sum_{i=1}^{c}\left[\sum_{\boldsymbol{x}\in\mathcal{X}_i}(\boldsymbol{x}-\boldsymbol{m}_i)(\boldsymbol{x}-\boldsymbol{m}_i)^\top\right] \tag{4}$$

$$\Sigma_B = \frac{1}{n}\sum_{i=1}^{c} n_i(\boldsymbol{m}_i-\boldsymbol{m})(\boldsymbol{m}_i-\boldsymbol{m})^\top \tag{5}$$

where c denotes the number of classes, and $\mathcal{X}_i$ and n_i indicate the data set and size of class i, respectively. Here, $\boldsymbol{m}_i$ and $\boldsymbol{m}$ indicate the mean vectors of $\mathcal{X}_i$ and the overall data, respectively. In this case, the number of classes c was set to three, i.e., minimum-jerk, constant-speed, and minus one-third power law.

For comparison, we also evaluated two networks that learned from either the minimum-jerk trajectories only (one pattern) or the minimum-jerk and constant-speed trajectories (two patterns). These networks trained with only one or two patterns correspond to subjects who failed at motor learning. Again, five trajectories of each velocity pattern were used in each case. In the evaluation, both networks were fed the trajectories of all three velocity patterns.

V. EXPERIMENTAL RESULTS

A. Experiment 1: Acquisition of Motor Representation

Fig. 4 shows the mean error in each network condition. The left and right bars denote the results of the RNNs and FNNs, respectively. The black, gray, and white bars denote the errors for the trajectories complying with minimum jerk, one-third power law, and constant speed, respectively. One-way ANOVAs (error × test pattern: minimum jerk, one-third power law, and constant speed) indicate significant main effects of the test patterns when the RNN learned prediction ($p < 0.01$) and when the RNN ($p < 0.05$) or FNN ($p < 0.01$) learned auto-association. *Post hoc* tests (Bonferroni) for predictive learning using the RNN [left bars in Fig. 4(a)] revealed that there are no significant differences between the errors for the minimum-jerk trajectories and the errors for the one-third power law trajectories ($p = 0.90$). It was also shown that the errors for the minimum-jerk or one-third power law trajectories are significantly smaller than the constant speed errors (either $p < 0.01$). Therefore, the acquired representation of this RNN includes the one-third power law, which enabled the model to distinguish the biological-like trajectories from those with constant speed.

In contrast, the output errors of the neural networks that learned auto-association could not be a biological criterion because the errors for the minimum-jerk trajectories were the largest [see Fig. 4(b)]. This was caused by the difference of the absolute amplitudes of the test patterns: the maximum input for the minimum-jerk, one-third power law, and constant-speed conditions were 0.81, 0.54, and 0.48, respectively. The models only acquired the auto-association, so the errors were more sensitive to the input magnitude than the trajectory smoothness.

Fig. 5 shows a distribution of the prediction error when an unlearned constant-speed trajectory was given to an RNN trained with minimum-jerk trajectories. The points are the positions of the inputs, and color indicates the value of the squared error at that time point. This clearly shows that the errors at larger curvatures are larger.

We analyzed the acquired representation of the neurons in the fourth layer with respect to acceleration which was computed by (2). Note that the acceleration information was not used for learning but computed only for the analysis in order to

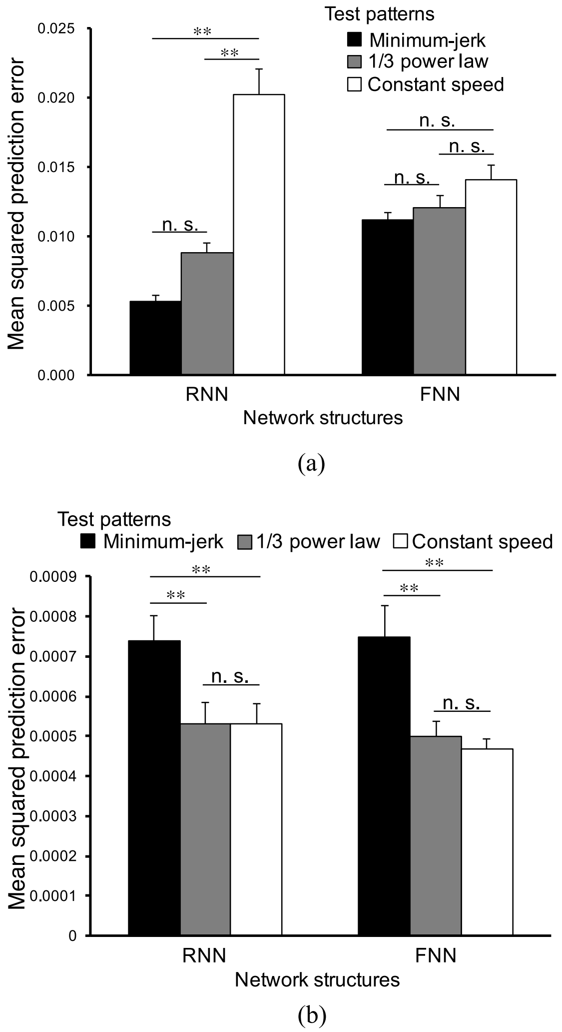


Fig. 4. Mean squared errors of the networks after learning (a) prediction or (b) auto-association of the desired trajectories. The black, gray, and white bars denote the errors for the minimum-jerk, one-third power law, and constant-speed trajectories, respectively. The left and right bars denote the errors output from the RNNs and FNNs, respectively. Stars indicate a significant difference with the solid line (**: $p < 0.01$, n.s.: $p > 0.05$). Error bars denote the standard error of the mean.

examine whether the network internally acquired the information through learning. Fig. 6 shows the correlation coefficients between activations of the neurons (rows) and population-coded acceleration of the inputs (columns) when the learned minimum-jerk trajectories are fed into the trained network. The figures indicate that these neurons acquired selectivity for a particular direction of acceleration. The correlation when an RNN learned motor prediction [Fig. 6 (a)] is clearer than for those in other conditions [Fig. 6(b), (c), and (d)]. The correlation was not very strong (the maximum coefficient was 0.34), suggesting that the modest representation of the acceleration allowed the model to represent not only the minimum-jerk property but also the one-third power law.

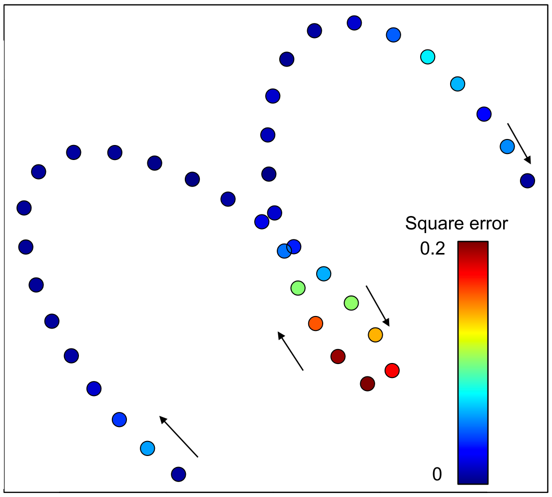
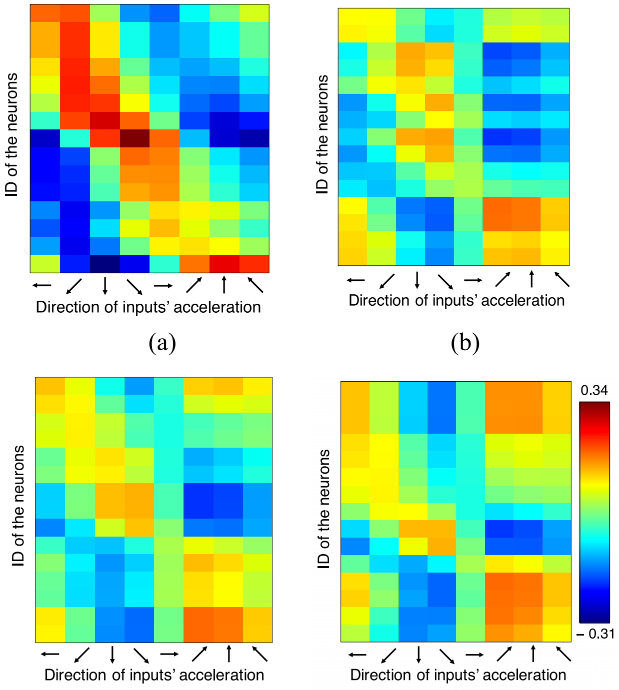


Fig. 5. Squared error distribution when an unknown trajectory with constant speed is fed to an RNN trained to predict minimum-jerk trajectories. The color of the points at regular time intervals denotes the magnitudes of the prediction error.



(c) (d)

Fig. 6. Correlation coefficients for fourth-layer neuron activations (rows) and input minimum-jerk trajectory acceleration (columns). The acceleration was coded as a neural population with direction selectivity. The networks of (a) and (b) had a recurrent structure while the networks of (c) and (d) were feedforward types. In (a) and (c), the networks learned motor prediction, but the networks in (b) and (d) learned auto-association.

B. Experiment 2: Learning From Acceleration and Velocity

Fig. 7 shows the prediction errors estimated by the RNN (left bars) or the FNN (right bars). Both networks were trained using the velocities and accelerations of the minimum-jerk trajectories. The overall errors were much lower than those in the former experiment, which used only velocities. There were also significant differences between conditions. One-way ANOVAs indicated significant main effects of the test patterns under both network structures (either $p < 0.01$). *Post hoc* tests (Bonferroni) revealed significant differences between the

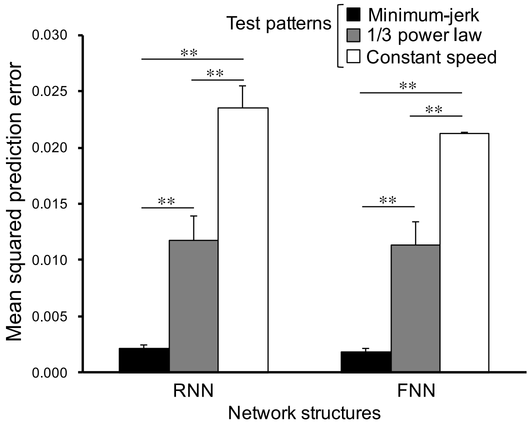


Fig. 7. Mean squared prediction errors for the RNNs (left bars) and FNNs (right bars) trained using the velocities and accelerations of the minimum-jerk trajectories. Stars indicate a significant difference (**: $p < 0.01$). Error bars denote the standard error of the mean.

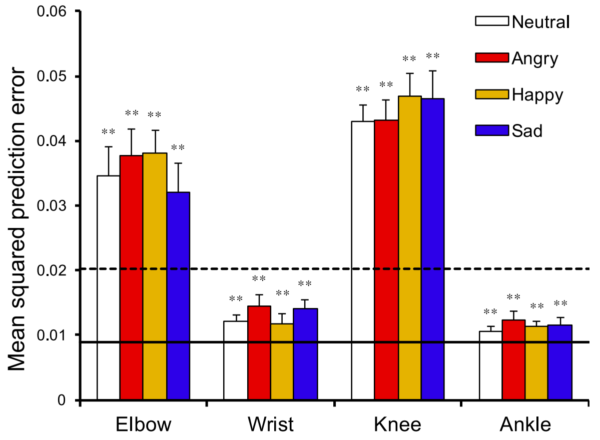


Fig. 8. Mean squared prediction errors for the trajectories of four body parts (elbow, wrist, knee, and ankle) produced by the RNN, which learned to predict minimum-jerk trajectories in experiment 1. The bar colors denote the emotional state of the walkers. The solid and broken lines show the mean errors for the one-third power law and constant-speed trajectories, respectively. Stars indicate a significant difference from the broken line (*: $p < 0.05$, **: $p < 0.01$). Error bars denote the standard error of the mean.

minimum-jerk and one-third power law trajectories, as well as between the one-third power law and constant-speed trajectories, regardless of network structure (both $p < 0.01$). Therefore, acceleration, as input, enabled the model to predict the minimum-jerk trajectories more accurately, even if the network did not have a recurrent structure. However, these networks over-fitted to the minimum-jerk trajectories, i.e., acceleration sequences with small change. Therefore, they did not generalize their internal representations to the one-third power law which does not have such smooth acceleration.

C. Experiment 3: Recognition of Human Walking PLDs

Fig. 8 shows the prediction errors for four body trajectories, averaged over individuals, that were produced by the RNN trained to predict minimum-jerk trajectories. The learning conditions were exactly the same as those in the first experiment [the basic results have already been shown as black bars in Figs. 4(a) and 6(a)]. We employed this network because it successfully represented biological motion (i.e., not only the minimum jerk but also the one-third power law). The colors of the bars in Fig. 8 represent the walkers' emotional states: white, red, yellow, and blue denote neutral, angry, happy, and sad, respectively. The stars on the top of the bars show the significant differences with respect to the mean prediction error for the constant-speed trajectories shown in the first experiment (broken line). The solid line denotes the mean prediction error for the one-third power law trajectories. All averaged errors for the trajectories of the wrist and ankle were significantly lower than the ones for the constant-speed trajectories (broken line) (all $p < 0.01$), and some of them did not significantly differ from the one-third power law trajectories. Therefore, the model could judge them as biological motions that were similar to the one-third power law trajectories. Our model did not recognize the trajectories of the elbow and knee as biological motion because these trajectories did not follow the one-third power law, i.e., they changed their velocities during motions with very small curvatures [see Fig. 3(a) and (c)]. A 4 (emotional state: neutral, angry, happy, sad) × 4 (body part: elbow, wrist, knee, ankle) × 2 (sex: male, female) × 28 (individual) mixed ANOVA revealed significant main effects for body parts ($p < 0.01$) and individuals ($p < 0.05$). Therefore, the model's performance was insensitive to walkers' emotional states or sex, although it may detect individual differences.

D. Experiment 4: Discrimination of Input Patterns

Fig. 9 shows the class separation index (3) of the activation of the third-layer neurons given trajectories in three classes of patterns (i.e., the minimum-jerk, constant speed, and minus one-third power law). Before the separation test, the networks learned the minimum-jerk trajectories only (left four bars), the minimum-jerk and constant-speed trajectories (middle four bars), or the minimum-jerk, constant-speed, and minus one-third power law trajectories (right four bars). The black, white, gray, and hatched bars, respectively, indicate the separation values in the cases of the RNN (prediction), FNN (prediction), RNN (auto-association), and FNN (auto-association). Clearly, the index is highest only when the RNN learned to predict all patterns.

VI. DISCUSSION

A. Predictive Learning in the PMd

This paper presented a PMd model that discriminates biological motions based on prediction errors. These errors were estimated by a neural network that represented self-induced limb movements, i.e., the one-third power law. The result of the first experiment proved the importance of a recurrent neural circuit to support predictive learning (see Fig. 4). The network internally yielded acceleration-like information (see Fig. 6) to predict the next input, resulting in a representation of a law that involves an acceleration function. Thus, the model could predict unknown trajectories according to the one-third power law, even when it was trained with minimum-jerk trajectories. There is a derivative gap between acceleration (the one-third

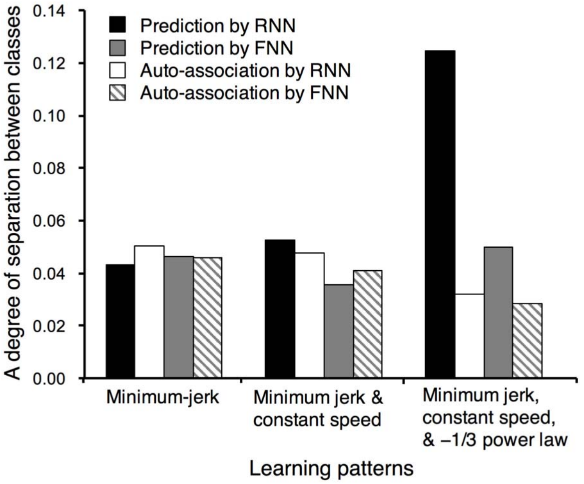


Fig. 9. Degree of separation between the classes of the velocity patterns in the neural activations in the third layer. The three bottom labels indicate the learning patterns fed to the network. The colors of the bars indicate the network conditions.

power law) and jerk (minimum-jerk law). The RNN implicitly represented the acceleration and jerk of the velocity inputs, and therefore, it had tolerance for the underlying laws. Thus, the representation overcame the derivative gap and extended to real biological movements (PLDs), as shown in Fig. 8.

We employed the minimum-jerk rule to generate body movements and assumed that the PMd represent the one-third power law. Some studies have asserted that the law emerges from the constraints of a body's musculoskeletal system rather than a constraint of the motor planning [54]–[56]. Gribble and Ostry [54] found that the viscoelasticity of muscles enables the effector's trajectories to satisfy the one-third power law even when the motor commands linearly change. To model a brain-body interaction from which the one-third power law emerges is an interesting future issue.

The second experiment demonstrated that a velocity was a sufficient input to represent biological motion. This assumption regarding the input is based on the neurophysiological findings [30], [52]. Intuitively, using accelerations as the input facilitates the acquisition of the one-third power law because it needs accelerations. However, our simulation showed that the acceleration inputs caused over-fitting to the minimum-jerk trajectories and prevented the network from generalizing to the one-third power law trajectories (see Fig. 7). This result again suggests the importance of an internal production of acceleration-like information by predictive learning. To our knowledge, the neural populations that explicitly code the accelerations of desired trajectories in the PMd have not been found. This may be a reason that neurons in this region improve their generalization capacity for predictive functions.

These results, which show that prediction is useful for biological motion detection, are consistent with neuroscientific studies on the PMd. This area was reported to activate during the prediction of an observed motion [27]–[29]. The PMd also treats the execution and imagination of self-induced limb movements as well as others' movements [22], [26], [28]. These studies imply that prediction, regardless of modality, is based on the representation of self-induced limb movements. Our model study can bridge the predictive nature of the PMd with the ability to detect biological motion (e.g., [8] and [15]).

This idea is supported by an fMRI study suggesting that prediction errors may lead to the uncanny valley (subjective strangeness or unfamiliarity) [57], [58] of an android [59]. They found more activation in the intraparietal sulcus, which is interconnected with the premotor area, when the subjects observed an android than a human or a mechanical robot. This android had a biological appearance but mechanically moved. The discrepancy between the appearance and movements causes the observer to generate prediction errors for its movements, leading to an uncanny impression. This seems consistent with our hypothesis that we perceive movements with large prediction errors as nonbiological movements.

In our model, the predictive learning from motor information alone can achieve the ability to discriminate biological motions. The conventional theory known as predictive coding has stated that predicting sensory information is a main function of the brain (e.g., [60] and [61]). A recent study reported that even a six-month-old infant's brain reacts to prediction errors between audiovisual associations [62]. A computational study proposed that sensorimotor predictive learning is a fundamental mechanism of cognitive development, e.g., imitation and helping behaviors [63]. In contrast to these theories, we emphasize the importance of the predictive learning based on self-induced motor information.

B. *Hierarchy of Biological Motion Perception*

In the third experiment, the model successfully judged the wrist and ankle trajectories of the PLD as biological motions (see Fig. 8). This result corresponds to a psychological study in which observers can recognize the trajectories of a walkers' body parts as biological [64]. Note that the performance of our model did not depend on the walkers' sex and emotional state. The model was able to robustly evaluate the PLD as biological motions despite differences based on sex and emotional state.

A number of studies have indicated that adults can recognize some properties of PLDs, e.g., the sex [65] and emotional state [66] of a walker. Our result implies that the model differs from a system for the identification of these properties. Troje [67] proposed a hierarchical model for the recognition of biological motion based on a number of psychological experiments. It mainly consists of two layers: 1) a higher system for the pattern recognition of action type, sex, and emotional state from global motion information and 2) a lower system for biological motion detection from a local trajectory. Our model corresponds to the lower system in this model, partially supporting Troje's hypothesis.

However, our ANOVA analysis clarified that our model had significant sensitivity to the individual differences of walkers (see Section V-C). This is because it was sensitive to trajectory jerk and some subjects exhibited jumping-like gaits. In addition, the model was not able to recognize the movements of an elbow and knee as biological motion because they did not comply with the one-third power law (see Fig. 8). These trajectories are intricately constrained by connected body parts and gravity. This is an example that shows that not all human movements satisfy the law because they are influenced by a body and the environment [56]. A psychological study is expected to investigate how an observer perceives the trajectory of such gaits and trajectories.

C. Improvement of Classification by Motor-Based Learning

The final experiment showed that motor-based learning from trajectories with three velocity patterns (the minimum jerk, constant speed, and minus one-third power law) improved discrimination performance between the patterns in the compressive layer (see Fig. 9). This was achieved only when an RNN learned motor prediction. The network classified the given trajectories into patterns in the compressive layer and switched the subsequent prediction strategy. The classification for prediction may be applied to discrimination of the visual stimuli. This computation could explain the experimental result reported by Beets *et al*. [31].

In our model, even given trajectories that followed the minus one-third power law, the prediction errors were small, indicating that they are similar motions as self-induced ones. It is reasonable, for example, that an athlete who is able to execute special movements recognizes the same movements, when he/she observes them, as natural or biological motion.

D. Neonates' Ability to Detect Biological Motion

While neonates' ability to detect biological motions has been regarded as innate [17], [18], [41], our model proposes that it can be acquired by motor experiences. This developmental view is supported by a report that the atypical kinematics profiles (e.g., greater jerkiness) of arm movements of subjects with autism spectrum disorder correlate with their low sensitivity to minimum-jerk trajectories [68]. This model can provide a computational basis for these developmental phenomena.

However, a previous study showed that neonates are able to distinguish standard PLDs from scrambled ones, i.e., ones in which each dot is initially positioned in a random manner [18]. This cannot be explained by our model because of its lack of a global perspective. It is expected that learning from the movements of an elbow and a shoulder as well as the movements of a hand allows the model to represent the link structure of the arm and detect a link structure of biological motion. In addition, neonates prefer upright biological motions to inverted ones [17], [18], but the model cannot explain this inversion effect because of a lack of gravity perception. Neonates, in contrast, have a nearly mature vestibular system that senses gravity [69]. Desired trajectories generate smooth trajectories that are optimized in an environment with gravity, which may enable the model to exhibit this effect.

In addition to biological motion detection, neonates are also reported to exhibit preference for face-like visual stimuli (e.g., [70] and [71]). There are computational models to explain this preference: the integration between the proprioceptive information of an arm and tactile information of the hand and face during facial double-touch allows them to construct their own facial patterns [72], [73]. Based on these models, the body representations acquired through motor experiences during the fetal period may shape postnatal capabilities to visually find biological motions and face-like patterns.

E. Mirror Property and Correspondence Problem

Our model can explain an ability to detect biological motion using a mirror property in the PMd. Mirror neurons discharge during both the execution of an action and the observation of another person performing the same action [74], which are typically observed in the ventral premotor area. This activation on the observation of another's actions is thought of as an understanding of the actions based on representations of the observer's corresponding actions [75], [76]. Therefore, these premotor neurons seem to encode visuomotor representations regardless of the agents, which recalls the common coding theory [77], [78]. This hypothesis claims that perceptual and motor representations share the same format, the so-called common code. In the common code, the observed actions of others directly evoke a correspondent motor representation in the motor-related area. This is also known as the direct matching hypothesis (e.g., [79] and [80]). In our model, biological motion representation is acquired as the common code to evaluate motor and visual prediction errors.

For the emergence of the mirror property, the correspondence problem [81], [82] is an important issue to be solved: how does an observer correspond an observed movement to the observer's motor system in spite of the big differences between the visual and motor (the observer's muscle activation) information? To avoid this problem, we assume that an observed movement is coded as local motion information, i.e., a population-coded velocity of a trajectory, in the same manner as motor information. If global motion or configuration information is used, differences among the modalities (e.g., coordination systems or body parts) appear. The population coding of the velocities allows the model to input them to the same RNN, resulting in biological motion detection. However, how the same format is automatically acquired is still an open question.

F. Whole-Brain Processing System for Biological Motion Perception

Finally, we propose a whole-brain processing system for biological motion perception to clarify the position of our model (see Fig. 10). A major suggestion is to separate the processing into two parts: 1) pattern recognition of the motion in the occipital and temporal lobes and 2) biological motion detection in the frontal and parietal lobes. A number of studies have reported that the STS activates during observation of PLDs (e.g., [1]–[4]). Peuskens *et al*. [4] hypothesized an interactive processing among the STS, MT/medial superior temporal area (MST), and interior temporal gyrus (ITG). The MT/MST analyzes the complex motion patterns of visual stimuli and sends them to the STS and ITG. The ITG extracts a human-like structure from a group of these motion patterns and sends it to the STS and MT/MST. Based on the form of the information, the MT/MST refines the analysis of the motion patterns, and the STS identifies the action type.

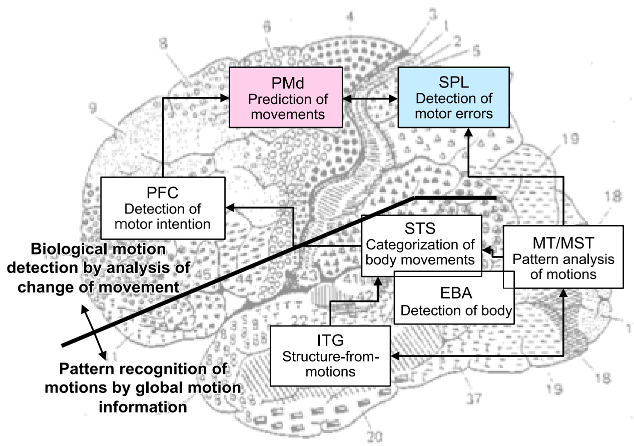


Fig. 10. Brain functions for biological motion perception. We hypothesize the two main functions: the occipital and temporal lobes address pattern recognition for body movements while the parietal and frontal lobes detect biological motions by predicting observed movements. The background brain image is in [83].

This system corresponds to the higher layer in Troje's hierarchical hypothesis [67] and was modeled by Giese and Poggio [33].

In addition, we hypothesize that a neural network of the PMd and SPL analyzes observed motions based on a representation of self-induced movements, as shown by the colored blocks in Fig. 10. When reaching for a moving object, for example, the PMd and SPL predict the velocity of the object and start to actuate the arm in anticipation. Therefore, it is reasonable that these areas predict observed movements and self-induced movements. Our simulation demonstrates that the motor prediction leads to the detection of biological motion. This system corresponds to the lower layer in Troje's hypothesis [67].

Further, the motor-related areas are modulated by top-down information from the prefrontal cortex (PFC) (e.g., [5], [84] and [85]). The PFC estimates an intention or mental state of an observed action, which is called mentalizing. Such top-down information is sent to the motor-related area and may be utilized to predict the target movement. Conventional studies for biological motion have focused on the occipital and temporal lobes. Further studies on the role of the parietal and frontal lobes in biological motion perception are expected.

VII. CONCLUSION

We hypothesized that the PMd predicts the velocity of an observed object based on the representation of desired smooth limb movements and evaluates how biological the observed motion is based on the prediction error. We hence conclude the following. First, the experiments revealed that predictive learning, achieved using an RNN, is essential for the acquisition of the biological motion representation, i.e., the one-third power law. Second, the learned RNN internally generated the acceleration information of the input velocity series and represented the relation between the curvature (computed by the acceleration) and speed. These results are consistent with the fact that the PMd originally functions as a motor predictor [25]–[29]. Third, the RNN also replicated the tendency that nonvisual motor learning to facilitate the discrimination of visual stimuli according to the learned law [31]. Finally, the RNN successfully discriminated the wrist and ankle trajectories of walking humans from nonbiological motions, irrespective of the walkers' sex and emotional state. This result is in line with the psychological theory that the ability to detect biological motion differs from the ability to recognize actor's sex and emotional state from motion information [67].

ACKNOWLEDGMENT

The authors would like to acknowledge the advice of Dr. C. Yoshida at Kyoto University.

REFERENCES

[1] E. Bonda, M. Petrides, D. Ostry, and A. Evans, "Specific involvement of human parietal systems and the amygdala in the perception of biological motion," *J. Neurosci.*, vol. 16, no. 11, pp. 3737–3744, 1996.

[2] R. J. Howard *et al.*, "A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain," *Current Biol.*, vol. 6, no. 8, pp. 1015–1019, 1996.

[3] E. Grossman *et al.*, "Brain areas involved in perception of biological motion," *J. Cogn. Neurosci.*, vol. 12, no. 5, pp. 711–720, 2000.

[4] H. Peuskens, J. Vanrie, K. Verfaillie, and G. A. Orban, "Specificity of regions processing biological motion," *Eur. J. Neurosci.*, vol. 21, no. 10, pp. 2864–2875, 2005.

[5] L. Centelles, C. Assaiante, B. Nazarian, J.-L. Anton, and C. Schmitz, "Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: A neuroimaging study," *PLoS One*, vol. 6, no. 1, 2011, Art. no. e15749.

[6] C. Di Dio *et al.*, "The neural correlates of velocity processing during the observation of a biological effector in the parietal and premotor cortex," *NeuroImage*, vol. 64, no. 1, pp. 425–436, 2013.

[7] E. Dayan *et al.*, "Neural representations of kinematic laws of motion: Evidence for action-perception coupling," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 51, pp. 20582–20587, 2007.

[8] A. Casile *et al.*, "Neuronal encoding of human kinematic invariants during action observation," *Cerebr. Cortex*, vol. 20, no. 7, pp. 1647–1655, 2010.

[9] F. Lacquaniti, C. Terzuolo, and P. Viviani, "The law relating the kinematic and figural aspects of drawing movements," *Acta Psychol.*, vol. 54, nos. 1–3, pp. 115–130, 1983.

[10] Y. Meirovitch, H. Harris, E. Dayan, A. Arieli, and T. Flash, "Alpha and beta band event-related desynchronization reflects kinematic regularities," *J. Neurosci.*, vol. 35, no. 4, pp. 1627–1637, 2015.

[11] C. de'Sperati and P. Viviani, "The relationship between curvature and velocity in two-dimensional smooth pursuit eye movements," *J. Neurosci.*, vol. 17, no. 10, pp. 3932–3945, 1997.

[12] Y. P. Ivanenko, R. Grasso, V. Macellari, and F. Lacquaniti, "Control of foot trajectory in human locomotion: Role of ground contact forces in simulated reduced gravity," *J. Neurophysiol.*, vol. 87, no. 6, pp. 3070–3089, 2002.

[13] G. Courtine *et al.*, "Performance of locomotion and foot grasping following a unilateral thoracic corticospinal tract lesion in monkeys (Macaca mulatta)," *Brain*, vol. 128, no. 10, pp. 2338–2358, 2005.

[14] M. Karklinsky and T. Flash, "Timing of continuous motor imagery: The two-thirds power law originates in trajectory planning," *J. Neurophysiol.*, vol. 113, no. 7, pp. 2490–2499, 2015.

[15] A. Harada and Y. Uno, "Generation and perception of curved trajectories in human arm movements," *Tech. Rep. IEICE NeuroComput.*, vol. 97, no. 623, pp. 237–244, 1998.

[16] K. Uemura, N. Fukumura, and Y. Uno, "Performance index of smoothness in generation and perception of curved movements of human arm," *Tech. Rep. IEICE NeuroComput.*, vol. 99, no. 684, pp. 71–78, 2000.

[17] F. Simion, L. Regolin, and H. Bulf, "A predisposition for biological motion in the newborn baby," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 2, pp. 809–813, 2008.

[18] L. Bardi, L. Regolin, and F. Simion, "Biological motion preference in humans at birth: Role of dynamic and configural properties," *Develop. Sci.*, vol. 14, no. 2, pp. 353–359, 2011.

[19] S. Zoia *et al.*, "Evidence of early development of action planning in the human foetus: A kinematic study," *Exp. Brain Res.*, vol. 176, no. 2, pp. 217–226, 2007.

[20] C. von Hofsten, "Structuring of early reaching movements: A longitudinal study," *J. Motor Behav.*, vol. 23, no. 4, pp. 280–292, 1991.

[21] J. C. Culham *et al.*, "Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas," *Exp. Brain Res.*, vol. 153, no. 2, pp. 180–189, 2003.

[22] F. Filimon, J. D. Nelson, D. J. Hagler, and M. I. Sereno, "Human cortical representations for reaching: Mirror neurons for execution, observation, and imagery," *Neuroimage*, vol. 37, no. 4, pp. 1315–1328, 2007.

[23] M. Matelli and G. Luppino, "Parietofrontal circuits for action and space perception in the macaque monkey," *Neuroimage*, vol. 14, no. 1, pp. 27–32, 2001.

[24] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, Eds., *Principles of Neural Science*, 5th ed. New York, NY, USA: McGraw-Hill, 2012.

[25] J. Diedrichsen, Y. Hashambhoy, T. Rane, and R. Shadmehr, "Neural correlates of reach errors," *J. Neurosci.*, vol. 25, no. 43, pp. 9919–9931, 2005.

[26] N. Malfait *et al.*, "fMRI activation during observation of others' reach errors," *J. Cogn. Neurosci.*, vol. 22, no. 7, pp. 1493–1503, 2010.

[27] W. Stadler *et al.*, "Predicting and memorizing observed action: Differential premotor cortex involvement," *Human Brain Mapping*, vol. 32, no. 5, pp. 677–687, 2011.

[28] U. Wolfensteller, R. I. Schubotz, and D. Y. Von Cramon, "Understanding non-biological dynamics with your own premotor system," *Neuroimage*, vol. 36, pp. T33–T43, Mar. 2007.

[29] R. I. Schubotz, "Prediction of external events with our motor system: Towards a new framework," *Trends Cogn. Sci.*, vol. 11, no. 5, pp. 211–218, 2007.

[30] R. Caminiti, P. B. Johnson, and A. Urbano, "Making arm movements within different parts of space: Dynamic aspects in the primate motor cortex," *J. Neurosci.*, vol. 10, no. 7, pp. 2039–2058, 1990.

[31] I. A. Beets, F. Rösler, and K. Fiehler, "Nonvisual motor learning improves visual motion perception: Evidence from violating the two-thirds power law," *J. Neurophysiol.*, vol. 104, no. 3, pp. 1612–1624, 2010.

[32] Y. Kawai, M. Asada, and Y. Nagai, "A model for biological motion detection based on motor prediction in the dorsal premotor area," in *Proc. 4th Joint IEEE Int. Conf. Develop. Learn. Epigenet. Robot.*, Genoa, Italy, 2014, pp. 249–255.

[33] M. A. Giese and T. Poggio, "Neural mechanisms for the recognition of biological movements," *Nat. Rev. Neurosci.*, vol. 4, no. 3, pp. 179–192, 2003.

[34] D. M. Wolpert, K. Doya, and M. Kawato, "A unifying computational framework for motor control and social interaction," *Philosoph. Trans. Roy. Soc. London B Biol. Sci.*, vol. 358, no. 1431, pp. 593–602, 2003.

[35] J. L. Copete, Y. Nagai, and M. Asada, "Motor development facilitates the prediction of others' actions through sensorimotor predictive learning," in *Proc. 4th Joint IEEE Int. Conf. Develop. Learn. Epigenet. Robot.*, Paris, France, 2016, pp. 223–229.

[36] G. Schillaci, V. V. Hafner, B. Lara, and M. Grosjean, "Is that me? Sensorimotor learning and self-other distinction in robotics," in *Proc. 8th ACM/IEEE Int. Conf. Human Robot Interact.*, Tokyo, Japan, 2013, pp. 223–224.

[37] F. Schrodt, G. Layher, H. Neumann, and M. V. Butz, "Embodied learning of a generative neural model for biological motion perception and inference," *Front. Comput. Neurosci.*, vol. 9, no. 79, 2015, doi: 10.3389/fncom.2015.00079.

[38] T. Sawaragi and T. Kudoh, "Self-reflective segmentation of human bodily motions using recurrent neural networks," *IEEE Trans. Ind. Electron.*, vol. 50, no. 5, pp. 903–911, Oct. 2003.

[39] A. N. Meltzoff and M. K. Moore, "Imitation of facial and manual gestures by human neonates," *Science*, vol. 198, no. 4312, pp. 75–78, 1977.

[40] A. N. Meltzoff and M. K. Moore, "Explaining facial imitation: A theoretical model," *Early Develop. Parenting*, vol. 6, nos. 3–4, pp. 179–192, 1997.

[41] D. Méary, E. Kitromilides, K. Mazens, C. Graff, and E. Gentaz, "Four-day-old human neonates look longer at non-biological motions of a single point-of-light," *PLoS One*, vol. 2, no. 1, pp. e186–e190, 2007.

[42] T. Flash and N. Hogan, "The coordination of arm movements: An experimentally confirmed mathematical model," *J. Neurosci.*, vol. 5, no. 7, pp. 1688–1703, 1985.

[43] P. Viviani and T. Flash, "Minimum-jerk, two-thirds power law, and isochrony: Converging approaches to movement planning," *J. Exp. Psychol. Human Perception Perform.*, vol. 21, no. 1, pp. 32–53, 1995.

[44] G. W. Cottrell, P. W. Munro, and D. Zipser, "Image compression by back propagation: A demonstration of extensional programming," in *Advances in Cognitive Science*, vol. 2, N. E. Sharkey, Ed. Chichester, U.K.: Ellis Horwood, 1988, pp. 208–241.

[45] B. Irie and M. Kawato, "Acquisition of internal representation by multi-layered perceptrons," *Inst. Electron. Inf. Commun. Eng.*, vol. 73, no. 8, pp. 1173–1178, 1990.

[46] R. Saegusa, H. Sakano, and S. Hashimoto, "Nonlinear principal component analysis to preserve the order of principal components," *Neurocomputing*, vol. 61, pp. 57–70, Oct. 2004.

[47] J. C. Houk, J. Keifer, and A. G. Barto, "Distributed motor commands in the limb premotor network," *Trends Neurosci.*, vol. 16, no. 1, pp. 27–33, 1993.

[48] P. F. Dominey, T. Inui, and M. Hoen, "Neural network processing of natural language: II. Towards a unified model of corticostriatal function in learning sentence comprehension and non-linguistic sequencing," *Brain Lang.*, vol. 109, nos. 2–3, pp. 80–92, 2009.

[49] R. Caminiti, P. B. Johnson, C. Galli, S. Ferraina, and Y. Burnod, "Making arm movements within different parts of space: The premotor and motor cortical representation of a coordinate system for reaching to visual targets," *J. Neurosci.*, vol. 11, no. 5, pp. 1182–1197, 1991.

[50] A. B. Schwartz, "Direct cortical representation of drawing," *Science*, vol. 265, no. 5171, pp. 540–542, 1994.

[51] A. B. Schwartz and D. W. Moran, "Motor cortical activity during drawing movements: Population representation during lemniscate tracing," *J. Neurophysiol.*, vol. 82, no. 5, pp. 2705–2718, 1999.

[52] J. H. Maunsell and D. C. Van Essen, "Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation," *J. Neurophysiol.*, vol. 49, no. 5, pp. 1127–1147, 1983.

[53] Y. Ma, H. M. Paterson, and F. E. Pollick, "A motion capture library for the study of identity, gender, and emotion perception from biological motion," *Behav. Res. Methods*, vol. 38, no. 1, pp. 134–141, 2006.

[54] P. L. Gribble and D. J. Ostry, "Origins of the power law relation between movement velocity and curvature: Modeling the effects of muscle mechanics and limb dynamics," *J. Neurophysiol.*, vol. 76, no. 5, pp. 2853–2860, 1996.

[55] A. Wing, C. Christou, and A. Waller, "Touching trajectories: The relation between speed curvature in exploring shape," in *Proc. Eurohaptics*, 2001, pp. 76–79.

[56] S. Schaal and D. Sternad, "Origins and violations of the 2/3 power law in rhythmic three-dimensional arm movements," *Exp. Brain Res.*, vol. 136, no. 1, pp. 60–72, 2001.

[57] M. Mori, "The uncanny valley," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.

[58] H. Ishiguro, "Android science: Conscious and subconscious recognition," *Connection Sci.*, vol. 18, no. 4, pp. 319–332, 2006.

[59] A. P. Saygin, T. Chaminade, H. Ishiguro, J. Driver, and C. Frith, "The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions," *Soc. Cogn. Affect. Neurosci.*, vol. 7, no. 4, pp. 413–422, 2012.

[60] K. Friston and S. Kiebel, "Predictive coding under the free-energy principle," *Philosoph. Trans. Roy. Soc. London B Biol. Sci.*, vol. 364, no. 1521, pp. 1211–1221, 2009.

[61] A. Clark, "Whatever next? Predictive brains, situated agents, and the future of cognitive science," *Behav. Brain Sci.*, vol. 36, no. 3, pp. 181–204, 2013.

[62] L. L. Emberson, J. E. Richards, and R. N. Aslin, "Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 31, pp. 9585–9590, 2015.

[63] Y. Nagai and M. Asada, "Predictive learning of sensorimotor information as a key for cognitive development," in *Proc. IROS Workshop Sensorimotor Contingencies Robot.*, 2015.

[64] D. H. Chang and N. F. Troje, "Perception of animacy and direction from local biological motion signals," *J. Vis.*, vol. 8, no. 5, pp. 1–10, 2008.

[65] C. D. Barclay, J. E. Cutting, and L. T. Kozlowski, "Temporal and spatial factors in gait perception that influence gender recognition," *Perception Psychophys.*, vol. 23, no. 2, pp. 145–152, 1978.

[66] A. P. Atkinson, W. H. Dittrich, A. J. Gemmell, and A. W. Young, "Emotion perception from dynamic and static body expressions in point-light and full-light displays," *Perception*, vol. 33, no. 6, pp. 717–746, 2004.

[67] N. F. Troje, "Biological motion perception," in *The Senses: A Comprehensive Reference*, vol. 2. San Diego, CA, USA: Academic Press, 2008, pp. 231–238.

[68] J. L. Cook, S.-J. Blakemore, and C. Press, "Atypical basic movement kinematics in autism spectrum conditions," *Brain*, vol. 136, no. 9, pp. 2816–2824, 2013.

[69] V. S. Dayal, J. Farkashidy, and A. Kokshanian, "Embryology of the ear," *Can. J. Otolaryngol.*, vol. 2, pp. 136–142, 1973.

[70] C. C. Goren, M. Sarty, and P. Y. Wu, "Visual following and pattern discrimination of face-like stimuli by newborn infants," *Pediatrics*, vol. 56, no. 4, pp. 544–549, 1975.

[71] M. H. Johnson, S. Dziurawiec, H. Ellis, and J. Morton, "Newborns' preferential tracking of face-like stimuli and its subsequent decline," *Cognition*, vol. 40, nos. 1–2, pp. 1–19, 1991.

[72] W. Song, T. Inui, and N. Takemura, "A neural network model for image formation of the face through double-touch without visual information," in *Proc. 2nd Int. Conf. Cogn. Neurodyn.*, Hangzhou, China, 2009, pp. 441–445.

[73] S. Fuke, M. Ogino, and M. Asada, "Body image constructed from motor and tactile images with visual information," *Int. J. Humanoid Robot.*, vol. 4, no. 2, pp. 347–364, 2007.

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

[75] V. Gallese and A. Goldman, "Mirror neurons and the simulation theory of mind-reading," *Trends Cogn. Sci.*, vol. 2, no. 12, pp. 493–501, 1998.

[76] V. Gallese, C. Keysers, and G. Rizzolatti, "A unifying view of the basis of social cognition," *Trends Cogn. Sci.*, vol. 8, no. 9, pp. 396–403, 2004.

[77] W. Prinz, *A Common Coding Approach to Perception and Action*. Heidelberg, Germany: Springer, 1990.

[78] W. Prinz, "Perception and action planning," *Eur. J. Cogn. Psychol.*, vol. 9, no. 2, pp. 129–154, 1997.

[79] G. Di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti, "Understanding motor events: A neurophysiological study," *Exp. Brain Res.*, vol. 91, no. 1, pp. 176–180, 1992.

[80] L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti, "Visuomotor neurons: Ambiguity of the discharge or 'motor' perception?" *Int. J. Psychophysiol.*, vol. 35, nos. 2–3, pp. 165–177, 2000.

[81] C. L. Nehaniv and K. Dautenhahn, "The correspondence problem," in *Imitation in Animals and Artifacts*. Cambridge, MA, USA: MIT Press, 2002, pp. 41–62.

[82] M. Brass and C. Heyes, "Imitation: Is cognitive neuroscience solving the correspondence problem?" *Trends Cogn. Sci.*, vol. 9, no. 10, pp. 489–495, 2005.

[83] K. Brodmann, *Vergleichende Lokalisationslehre der Grobhirnrinde*. Leipzig, Germany: Barth, 1909.

[84] J. Stanley, E. Gowen, and R. C. Miall, "How instructions modify perception: An fMRI study investigating brain areas involved in attributing human agency," *Neuroimage*, vol. 52, no. 1, pp. 389–400, 2010.

[85] Y. Wang and A. F. C. Hamilton, "Social top-down response modulation (STORM): A model of the control of mimicry in social interaction," *Front. Human Neurosci.*, vol. 6, no. 153, 2012, doi: 10.3389/fnhum.2012.00153.

Yuji Kawai received the master's and Ph.D. degrees in engineering from Osaka University, Suita, Japan, in 2013 and 2017, respectively.

He has been a Specially Appointed Assistant Professor with the Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University, since 2016. His current research interests include constructive understanding of human cognitive development and computational neuroscience.

Yukie Nagai (M'09) received the master's degree from Aoyama Gakuin University, Tokyo, Japan, in 1999, and the Ph.D. degree from Osaka University, Osaka, Japan, in 2004, both in engineering.

She was a Post-Doctoral Researcher with the National Institute of Information and Communications Technology, Kyoto, Japan, from 2004 to 2006, and with Bielefeld University, Bielefeld, Germany, from 2006 to 2009, where she was also with the Research Institute for Cognition and Robotics. She has then been a Specially Appointed Associate Professor with Osaka University since 2009. Her current research interests include computational modeling of human cognitive functions such as self-other cognition, imitation, and joint attention, and design of assistant systems for developmental disorders. Since 2012, she has been the Project Leader with MEXT Grant-in-Aid for Scientific Research on Innovative Areas entitled *Computational Modeling of Social Cognitive Development and Design of Assistance Systems for Developmental Disorders*. Since 2016, she has also been the Project Leader of JST CREST entitled *Cognitive Mirroring: Assisting People With Developmental Disorders by Means of Self-Understanding and Social Sharing of Cognitive Processes*.

Minoru Asada (M'88–SM'02–F'05) received the B.E., M.E., and Ph.D. degrees in control engineering from Osaka University, Osaka, Japan, in 1977, 1979, and 1982, respectively.

He is a Professor with Osaka University in 1995, where he has been a Professor with the Department of Adaptive Machine Systems, Graduate School of Engineering, since 1997.

Dr. Asada was a recipient of many awards such as the Best Paper Award at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '92) and the Commendation by the Minister of Education, Culture, Sports, Science and Technology, Japanese Government as Person of Distinguished Services to Enlightening People on Science and Technology. He is one of the founders of RoboCup, and the Former President of the International RoboCup Federation from 2002 to 2008. Since 2005, he has been the Research Director of ASADA Synergistic Intelligence Project at Exploratory Research for Advanced Technology by Japan Science and Technology Agency. He is currently a Principal Investigator of Grants-in-Aid for Scientific Research entitled *Constructive Developmental Science Based on Understanding the Process From Neurodynamics to Social Interaction*.