Cognitive Developmental Robotics: A Survey

Minoru Asada, *Fellow, IEEE*, Koh Hosoda, *Member, IEEE*, Yasuo Kuniyoshi, *Member, IEEE*,
Hiroshi Ishiguro, *Member, IEEE*, Toshio Inui, Yuichiro Yoshikawa, Masaki Ogino, and Chisato Yoshida

***Abstract*—Cognitive developmental robotics (CDR) aims to provide new understanding of how human's higher cognitive functions develop by means of a synthetic approach that developmentally constructs cognitive functions. The core idea of CDR is "physical embodiment" that enables information structuring through interactions with the environment, including other agents. The idea is shaped based on the hypothesized development model of human cognitive functions from body representation to social behavior. Along with the model, studies of CDR and related works are introduced, and discussion on the model and future issues are argued.**

***Index Terms*—Cognitive developmental robotics (CDR), development model, synthetic approach.**

I. INTRODUCTION

EMERGENCE of higher order cognitive functions through learning and development is one of the greatest challenges in trying to make artificial systems more intelligent since existing systems are of limited capability even in fixed environments. Related disciplines are not just artificial intelligence and robotics but also neuroscience, cognitive science, developmental psychology, sociology, and so on, and we share this challenge. An obvious fact is that we have insufficient knowledge and too superficial implementations based on such knowledge to declare that we have only one unique solution to the mystery. The main reasons are the following.

- There is little knowledge and few facts on the mechanism of higher order human cognitive functions; therefore, the artificial systems that aim at realizing such functions are based on the designers' shallow understanding of them.
- A more serious issue is how these functions are learned and/or developed from a viewpoint of design.
- Further, is the current understanding and realization of the primary functions sufficient if we suppose that the higher order cognitive functions are acquired through the development process from these primary functions?

One possibility to answer these claims and questions is to discuss how higher order cognitive functions are acquired involving the context and dynamics of the whole system instead of separately realizing each higher order function as a single module. A promising approach is a synthetic one based on both the explanation theory and, more importantly, the design theory that is expected to fill in the gap between the existing disciplines instead of staying in one closed discipline, and to provide new understanding of human cognitive development.

A key idea is "physical embodiment" whose meaning has been frequently defined and argued already (e.g., [1]–[8]). Kuniyoshi [9] described it as follows.

> The agent's physical body specifies the constraints on the interaction between the agent and its environment that generate the rich contents of its process or consequences. It also gives the meaningful structure to the interaction with environment, and is the physical infrastructure to form the cognition and action.

The key concept of the above "physical embodiment" is shaped in the context of development as follows. At the early stage of human development (embryo, fetus, neonate, infant, and so on), interactions with various physical environments have a major role in determining the information structuring inside the individual such as body representation, motor image, and object permanency. On the other hand, at the later stage, social behaviors such as early communication, joint attention, imitation of various actions including vocalization, empathy, and verbal communication gradually emerged due to interactions with other agents. Regardless of the premature or mature state of the individual, the common aspect of these developmental processes is a sort of "scaffolding" by the environment including other agents that triggers the sensorimotor mapping and promotes the infants' autonomy, adaptability, and sociality, directly or indirectly, and explicitly or implicitly.

A representative synthetic approach is cognitive developmental robotics (CDR) [5]. Similar approaches can be found in [7] or [10], but CDR puts more emphasis on the human/humanoid cognitive development. A slightly different approach is taken by the ATR team [11], which aims to program humanoid behavior through the observation and understanding of human behavior and vice versa. Though partially sharing the purpose of human understanding, they do not exactly deal with developmental aspect.

As mentioned above, the developmental process consists of two phases: the individual development at an early stage and the social development through interaction between individuals later on. The former relates mainly to neuroscience (internal

Manuscript received December 14, 2008; revised February 11, 2009. First published April 28, 2009; current version published May 29, 2009.

M. Asada, K. Hosoda, and H. Ishiguro are with the JST ERATO Asada Synergistic Intelligence Project, Osaka, Japan; and with the Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University, Osaka 565-0871, Japan.

Y. Kuniyoshi is with the JST ERATO Asada Synergistic Intelligence Project, Osaka, Japan; and the University of Tokyo, Tokyo, Japan.

T. Inui is with the JST ERATO Asada Synergistic Intelligence Project, Osaka, Japan; and Kyoto University, Kyoto, Japan.

Y. Yoshikawa, M. Ogino, and C. Yoshida are with the JST ERATO Asada Synergistic Intelligence Project, Osaka, Japan.

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/TAMD.2009.2021702

mechanism), and the latter to cognitive science and developmental psychology (behavior observation). Intrinsically, both should be seamless, but there is a big difference between them at the representation level for the research target to be understood. CDR aims not at simply filling the gap between them but more challenging at building a new paradigm that provides new understanding of ourselves and, at the same time, new design theory of humanoids symbiotic with us. So far, CDR has been mainly focusing on the computational model of cognitive development, but in order to more deeply understand how humans develop, robots can be used as a new means of reliable reproduction tools in certain situations such as psychological experiments. The following is a summary.

A) Construction of computational model of cognitive development:
   1) hypothesis generation: proposal of a computational model or hypothesis based on knowledge from existing disciplines;
   2) computer simulation: simulation of the process difficult to implement with real robots such as physical body growth;
   3) hypothesis verification with real agents (humans, animals, and robots), then go to 1).

B) Offer new means or data to better understand human developmental process → mutual feedback with A):
   1) measurement of brain activity by imaging methods;
   2) verification using human subjects or animal ones;
   3) providing the robot as a reliable reproduction tool in (psychological) experiments.

This paper gives a survey of CDR starting from a brief overview of the various aspects of infant development that provide the fundamental knowledge and inspiration for CDR. Next, we introduce the model of development toward the exploration for the design principle of cognitive development based on the current knowledge of neuroscience and developmental psychology. The model starts from the fetal sensorimotor mapping in the womb and moves to the social behavior learning through body representation, motor skill development, and spatial perception. Along this model, the following sections give an overview of related studies.

1) The most fundamental structure for motions, that is, the spinal cord–brain stem–cortex network that includes the simulation of fetal sensorimotor development.
2) Mechanism of dynamic motions of whole body from rolling over and crawling to walking and also jumping (voluntary movements). This section focuses on the physical implementation of dynamic motions since the research platform is very important for CDR and related research disciplines. Pneumatic actuators are tested as artificial muscles to generate dynamic motions and to understand the mechanism of humans' dynamic motions.
3) Body/motor representation and spatial perception to link the individual development and a social one between individuals.
4) The developmental of social behaviors such as early communication, action execution and understanding, vocal imitation, joint attention, and empathy development, showing what are the key aspects to trigger each social behavior from a viewpoint of scaffolding by a caregiver.

Last, discussion and future issues are given. The references are not exhaustive but selected in order to focus on the issues of CDR[1].

II. VARIOUS ASPECTS OF DEVELOPMENT

A. *Normal Development of Fetus and Infant*

Recent imaging technology such as three-dimensional (3-D) ultrasound movies have enabled observation of the various kinds of fetal movements in the womb after several weeks of gestation and reveals the possibility of fetus learning in the womb [14]. Vries *et al.* [13] reported that fetal motility started from the early state of "just discern movements (7.5 weeks)" to the later state of "sucking and swallow (12.5–14.5 weeks)" through "startle, general movements, hiccup, isolated arm movements, isolated leg movements, head retroflexion, head rotation, hand/face contact, breathing movements, jaw opening, stretch, head anteflexion, and yawn." Campbell [15] also reported that the eyes of the fetus open around 26 weeks of gestation and that the fetus often touches its face with its hands during embryonic weeks 24 and 27.

Regarding the fetal development of sense, touch is the first to develop, and then other senses such as taste, auditory, and vision start to develop. Chamberlain stated as follows: just before 8 weeks gestational age, the first sensitivity to touch manifests in a set of protective movements to avoid a mere hair stroke on the cheek. From this early date, experiments with a hair stroke on various parts of the embryonic body show that skin sensitivity quickly extends to the genital area (10 weeks), palms (11 weeks), and soles (12 weeks). These areas of first sensitivity are those that will have the greatest number and variety of sensory receptors in adults. By 17 weeks, all parts of the abdomen and buttocks are sensitive. Skin is marvelously complex, containing a hundred varieties of cells that seem especially sensitive to heat, cold, pressure, and pain. By 32 weeks, nearly every part of the body is sensitive to the same light stroke of a single hair. Both hearing and vision start about 18 weeks after gestation and develop to complete their perception at around 25 weeks.

Moreover, it is reported that visual stimulation from the outside of the maternal body can activate the fetal brain [16]. Fig. 1 shows the emergence of fetal movements with the development of fetal senses reflecting the above knowledge.

After birth, infants are supposed to gradually develop body representation, categories for graspable objects, capability of mental simulation of actions, and so on through their learning processes. For example, hand regard at the fifth month means learning of the forward and inverse models of the hand. Table I shows typical behaviors and their corresponding targets to learn.

Thus, human fetuses and infants expose cognitive developmental process with remarkable vigor. However, the early cognitive development of the first year after the birth is difficult

---

[1] The JST ERATO Asada Synergistic Intelligence Project (http://www.jeap.org/) has been doing many studies on this topic.

[2] http://www.birthpsychology.com/lifebefore/fetalsense.html.

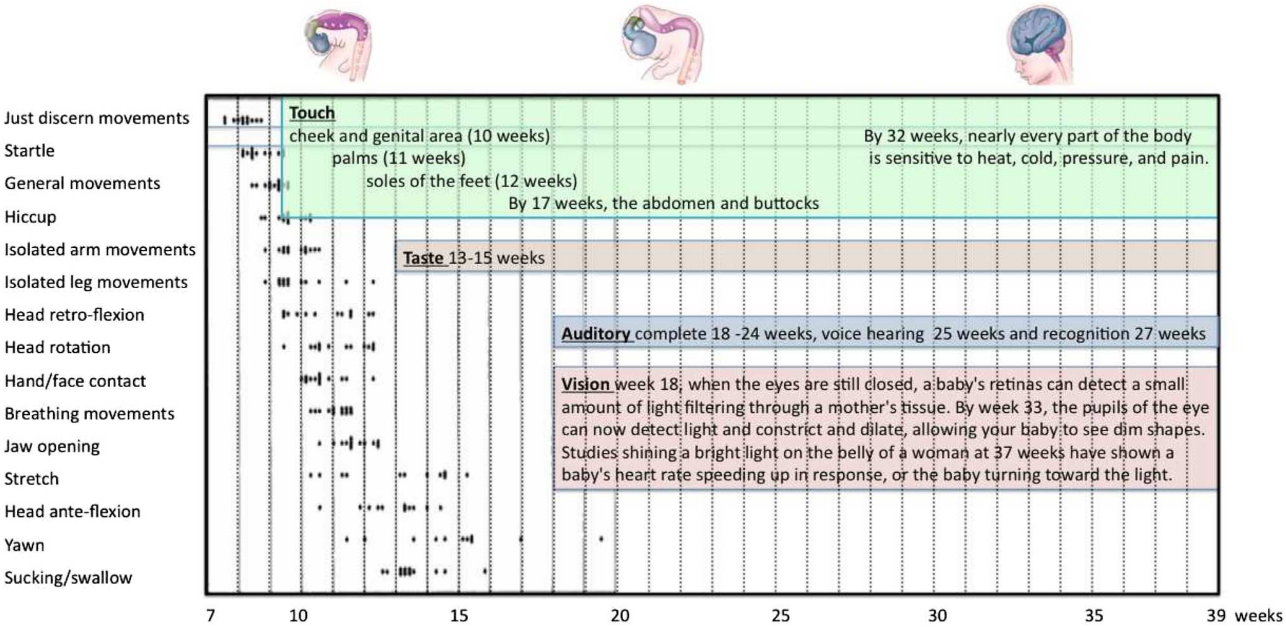


Fig. 1. Emergence of fetal movements and sense (brain figures on the top are adapted from [12, Fig. 22.5], emergence of movements is adapted from [13, Fig. 1], and fetal senses are adapted from [14]).

TABLE I
INFANT DEVELOPMENT AND LEARNING TARGETS

months	behaviors	learning targets
5	hand regard	forward and inverse models of the hand
6	finger the other's face	integration of visuo-tactile sensation of the face
	observe objects from different viewpoints	3-D object recognition
7	drop objects and observe the result	causality and permanency of objects
8	hit objects	dynamics model of objects
9	drum or bring a cup to mouth	tool use
10	imitate movements	imitation of unseen movements
11	fine grasp and carry objects to others	action recognition and generation, cooperation
12	pretend	mental simulation

to visualize since the imaging technology applicable to this age is still very limited, and the following points are suggested.

1) We cannot derive the infants' brain structure and functions from the adults' ones, nor should do it [17]–[19].
2) Brain regions for function development and function maintenance are not the same. During early language development, damage of the region in the right hemisphere is much more serious than that of the left [20].
3) The attention mechanism develops from the bottom-up ones, such as visual saliency map, to the top-down one needed to accomplish the specified task, and the related brain regions shift from posterior to anterior ones [21].
4) Even though the appearances of the performances look similar, their neural structures might be different. Generally, the shift from subcortical to cortical areas is observed from a macroscopic viewpoint. The brain region active for responding to joint attention is the same as the region of general attention (the left parietal lobe), but that for the ability to initiate joint attention includes the prefrontal area and close to the area for language [21], [22].

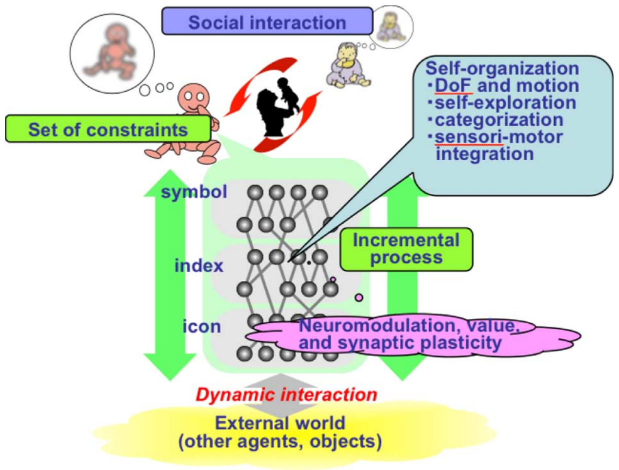


Fig. 2. Various aspects of the development from viewpoints of external observation, internal structure, its infrastructure, and social structure. Here, we briefly review the issue considering the underlying mechanisms in different forms.

B. Development From a Viewpoint of Synthetic Approaches

Here, we briefly review the facets of development in the survey by Lungarella *et al.* [23] from viewpoints of external observation, internal structure, its infrastructure, and social structure, especially focusing on the underlying mechanisms in different forms. Fig. 2 summarizes the various aspects of the development according to this review.

From the observation of the behaviors, the developmental process of infants can be regarded as one that is not centrally

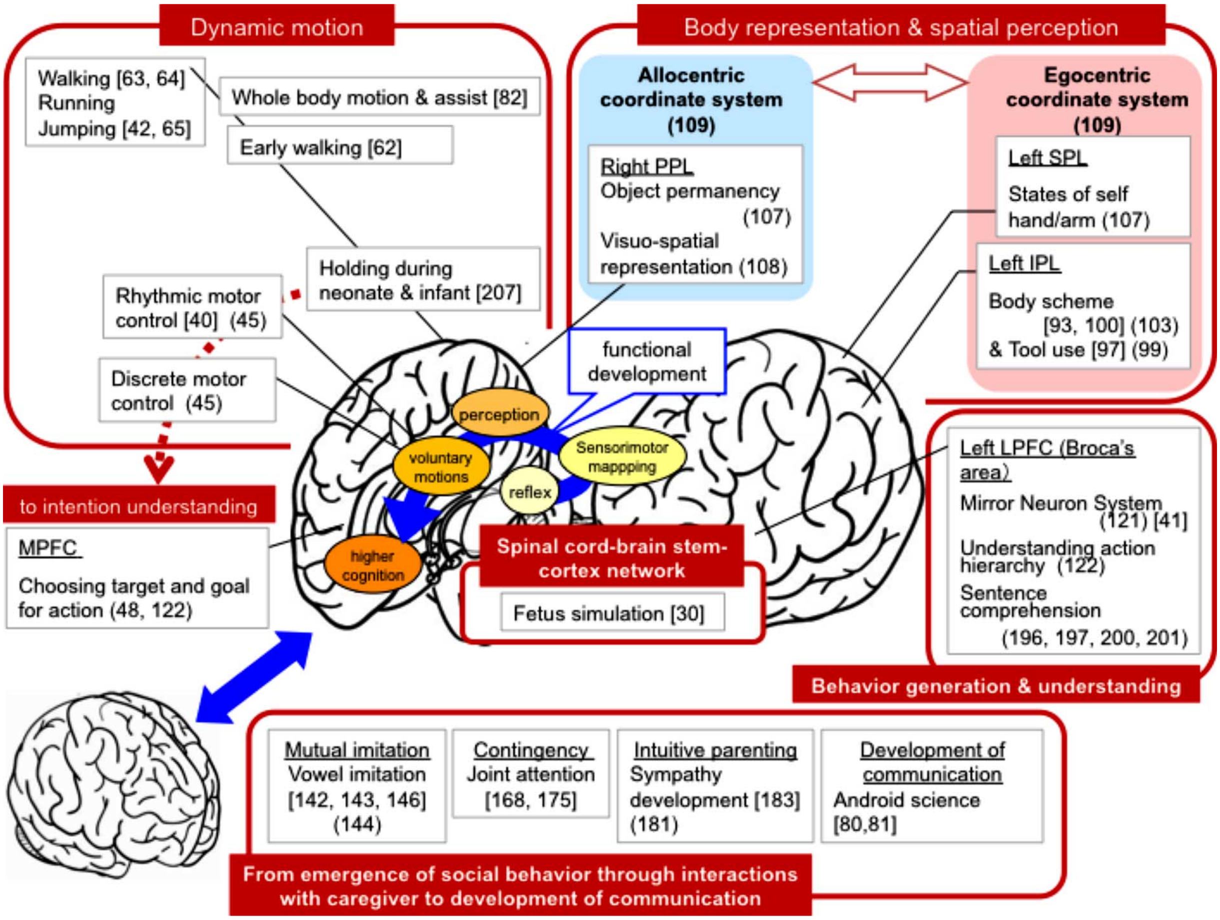


Fig. 3. The model of cognitive development that starts from the fetal sensorimotor mapping in the womb to the social behavior learning through body representation, motor skill development, and spatial perception. The numbers inside brackets ([]) and parentheses (()) indicate the references categorized into A (construction of computational models of cognitive development) and B (providing new means or data to better understand human developmental processes) in the Introduction, respectively.

controlled but instead a distributed and self-organized process. During the developmental stage, the later structure is constructed on top of the former structure that is a neither complete nor efficient behavior representation. This is one of the biggest differences from artificial systems [23]. Ecological constraints of infants are not always handicaps but can also serve to promote the development. The intrinsic tendency of coordination or pattern formation between brain, body, and environment is often referred to as entrainment, or intrinsic dynamics [24]. Self-exploration plays an important role in infancy, in that infants' "sense of the bodily self" to some extent emerges from a systematic exploration of the perceptual consequences of their self-produced actions [25], [26].

The consequence of active exploration and interaction with the environment is regarded as perceptual categorization and concept formation in developmental psychology. Sense and some sort of perception are processed independent of motion, but perceptual categorization depends on the interaction between sensory and motor systems. In the self-organization, some processes are regulated by neuromodulators that relate to value or synaptic plasticity, and there is a study to predict this kind of interaction from the computational model of metalearning [27].

Macroscopically, the quality of involvement with caregiver or others promotes the infants' autonomy, adaptability, and sociality. Scaffolding by a caregiver plays an important role in cognitive, social, and skill development. Infants have "sensitive periods" to caregivers' responses, and the caregivers regulate their responses to the infants.

C. *Model of Cognitive Development*

Let us consider the model of cognitive development based on the various aspects mentioned in the previous section. The major functional structure of the human brain–spine system is a hierarchical one reflecting the evolutionary process, and consists of spine, brain stem, diencephalon, cerebellum, limbic system, basal ganglia, and neocortex. Here, we regard this hierarchy as the first analogy toward the cognitive developmental model, and the flow of functional development is indicated at the center of Fig. 3, that is, reflex, sensorimotor mapping, perception, voluntary motion, and higher order cognition.

Hereafter, we briefly show the flow of the development model with studies related to CDR and the related disciplines, and discuss the validity of the model for cognitive development. The numbers in Fig. 3 inside brackets ([]) and parentheses (()) indicate the references cited in the following sections categorized into A (construction of computational models of cognitive development) and B (providing new means or data to better understand human developmental processes) in the Introduction, respectively.

III. SPINAL CORD–BRAIN STEM–CORTEX NETWORK

The hierarchical structure of motor control starts from spinal reflex without any control from the central nervous system (CNS) and generation of fixed motor patterns by medulla that coordinate the movements of body parts. Next is motion assembly by the CNS in terms of the fixed motor patterns, and sensorimotor integration by the parietal association area that leads to the representation and recognition of body and space. Then, the motor area in the cerebrum represents the repertoire of various kinds of motions and combines/switches/executes the motions in close cooperation with basal ganglia.

One of the research issues of CDR at this stage is acquisition of body representation and is the most fundamental issue related to cognitive development based on physical embodiment. How the body representations, called body schema or body image, are acquired is a big mystery. Neonatal imitation [28] in particular has been a hot topic causing a controversial argument between “innate” and “learned.” As we have mentioned from 3-D ultrasound imaging of the fetus movements, the fetuses start touch motions with their body parts such as face and arm at least 14 or 15 weeks after gestation. Among studies inspired by these findings, Kuniyoshi and Sangawa [29] have done a striking simulation of fetus development in the womb and emergence of neonatal behaviors. To the best of our knowledge, this is the first trial of the simulation that indicates how the fetus brain and body interact with each other in the womb.

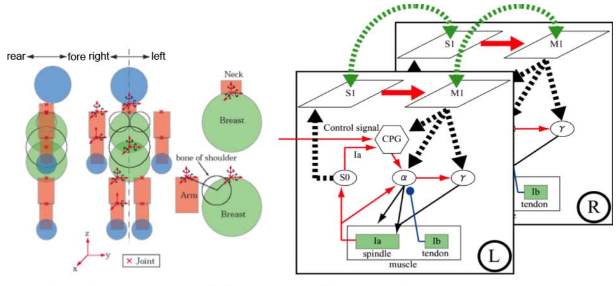
A. Emergence of Fetal and Neonatal Movements

Kuniyoshi and Sangawa [29] constructed a fetus simulation model and showed that various meaningful motor patterns emerge without “innate” motor primitives. The model consists of a musculoskeletal body floating in a uterus environment (elastic wall and liquid) and a minimal nervous system consisting of spine, medulla, and primary sensory/motor cortical areas.

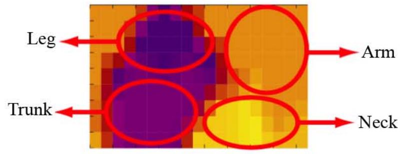
Besides the global connections depicted in Fig. 4(b), the only predefined (“innate”) circuits in the nervous system are 1) stretch reflex in the spinal circuit and 2) Bonhoffer–van der Pol (BVP) oscillator neurons in the medulla circuit, each connected to an individual muscle only. There is no predefined circuit specifying coordination of multiple muscles.

The BVP neurons have often been used as CPG units (e.g., [30] and [31] for biped walking, and [32] for quadruped walking). In such applications, the interconnections between CPG units are explicitly designed, with careful tuning of the parameters and/or external signals. However, the above fetus model assumes none of these, relying on multiple nonlinear oscillators to be coupled through embodiment.

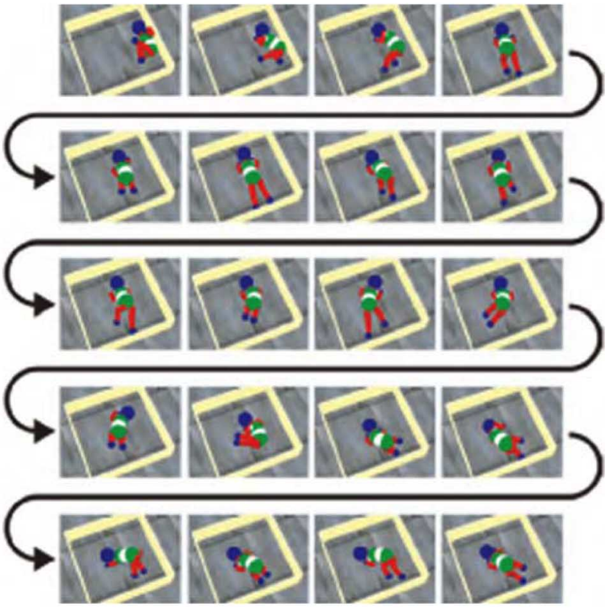
Therefore the observed whole-body motor patterns are purely emergent from the interaction between the body, environment, and the nervous system. This differs from the early pioneering work by Suzuki *et al.* [33] that formed motor patterns through the interaction between oscillators.



(a) a fetus body model (b) sensorimotor brain model



(c) mapping from M1 to α



(d) neonatal movement

Fig. 4. Fetal sensorimotor mapping and neonatal movements. (a) A fetus body model that consists of cylindrical or spherical body segments, connected to each other with constrained joints. (b) A brain model, lateral organization of the nervous system consisting of CPG (BVP neurons), S1 (primary somatosensory area), M1 (primary motor area), and so on. (c) The self-organized map from M1 to α motor neurons exhibiting separation into areas corresponding to different body parts. (d) The “neonate” model exhibits emergent motor behaviors such as rolling over and crawling-like motion.

Fig. 4(a) indicates the fetus body model that consists of cylindrical or spherical body segments, connected to each other with constrained joints. They defined 19 segments. Size, mass, and moment of inertia of each segment are determined to match the average fetus/neonate based on known data (e.g., [34] and [35]). Other detailed parameters such as joint angle limits, contraction force, cross-sectional area of the muscles are also determined or estimated from the literature such as [36] (see [29] for more references).

The fetus brain model is shown in Fig. 4(b), where lateral organization of the nervous system consisting of CPG (BVP neurons), S1 (primary somatosensory area), M1 (primary motor area), and so on. The fundamental structure is given, but connection weights are initially random (arrow and filled circle represent excitatory and inhibitory connections, respectively. Thick broken lines represent all to all connections with plasticity). Hebbian learning and self-organizing mapping method are used to determine the connection weights between brain parts. Fig. 4(c) shows the self-organized map from M1 to α motor neurons exhibiting separation into areas corresponding to different body parts. The "neonate" model exhibits emergent motor behaviors such as rolling over and crawling-like motion as shown in Fig. 4(d).

B. *Other Synthetic Approaches*

Chen and others [37] reported a self-organizing cortical map model for a simplified one-arm body. They propose to use the model for studying lesion effects, which suggests a potentially very important area of application in a larger scale for our model. Our cortical model is an extension of their model. However, their paper does not consider any possibility of emergence and development of behavior patterns from embodied interactions.

Recently, Izhikevich and Edelman [38] simulated a detailed large-scale thalamocortical model based on experimental measures in several mammalian species. Although they have shown interesting results of brain activity, the current system is still at the calibration stage; therefore no meaningful input reflecting physical embodiment is given. We suppose that the structured information coming from the physical embodiment and its interaction with the environment is crucial for determining the connection weights, and consequently the whole brain activity including the motor outputs. That is, body shapes brain [9].

C. *Future Issues for More Details and Verification*

Many issues yet to be addressed can be roughly classified into two types. The first one is the addition and refinement of the brain regions. Kinjo *et al.* [39] added the cerebellum model in order to account for the memory and reproduction of the emerged periodic motions, or *motor primitives*. Pitti *et al.* [40] proposed a model of cross modal learning of haptics and vision, showing how it self-organizes a mirror-system property after grasping experiences. These are a few examples of this case.

The second one is body parts including face, hand, and other sensor organs such as vision and auditory, and comparisons with experimental data on real agents (human infant and robots). These are now ongoing studies by research groups such as the JST ERATO Asada Project.

IV. MECHANISM OF DYNAMIC MOTIONS

A. *Development of Motor Skills*

In the previous section, Kuniyoshi and Sangawa [29] showed the emergence of the fetal and neonatal movements that do not seem conscious ones mainly regulated by spinal cord and brain stem. After birth, infants start to expose various kinds of whole body movements such as rolling over, crawling, sitting, standing with and without support, and walking with and without support. During such a developmental process, the movements change from unconscious ones to conscious rhythmic ones, then more complicated ones, and the related brain regions seem to extend from posterior regions (brain stem and cerebellum) to anterior ones (basal ganglia, cerebral cortex).

Righetti and Ijspeert designed a pattern generator network for rhythmic crawling motion of a baby humanoid model [42]. They recorded the trajectories of the limbs of real crawling babies and, based on the data, designed the oscillator and the network. They demonstrated that the model can reproduce almost the same motion by a dynamic simulation. To realize baby-like crawling, Degallier *et al.* developed an infant-like robot platform "iCub" [43]. In these studies, the trajectories of the model are directly generated by the oscillator, and the interaction between the body and environment has not been taken into account.

Regarding the relationship between rhythmic movements and discrete ones, Schaal *et al.* [44] showed that in addition to areas activated in rhythmic movement, discrete movement involves several higher cortical planning areas, even when both movement conditions are confined to the same single wrist joint. While many behavioral studies (e.g., [24], [45]) have focused on rhythmic models, subsuming discrete movement as a special case, neurophysiological and computational research on arm motor control (e.g., [46]) has focused almost exclusively on discrete movements, essentially assuming similar neural circuitry for rhythmic tasks. Schaal *et al.* provided neuroscientific evidence that rhythmic arm movement cannot be part of a more general discrete movement system and may require separate neurophysiological and theoretical treatment.

The neural mechanisms for motor control shown by Schaal *et al.* [44] were related to moving one's own arm in self-chosen comfortable frequency as well as triggered and maintained by cues. This finding suggested that the medial frontal cortex might be involved in adapting the timings to produce one's movement to rhythmic triggers existing in the external world. Through stronger connections with neural correlate for motor control, the regions in anterior cingulate cortex (ACC) that were more activated in Schaal *et al.* have a role of choosing the appropriate action based on the predicted consequences for possible alternatives ([47]). Extending and developing these regions to the anterior part of medial frontal cortex, it would correspond to the responsible region for "theory of mind," one of the social cognitive functions for estimating others' intentions, mental states, and contexts, as many brain imaging studies have shown. This extension is congruent with the developmental course of motor control and brain functions in infancy [48]. As for social cognition in infancy, there might be many functions that were acquired as behavioral prototypes and are overt as some kinds of voluntary movements. To be functional as the behavioral basis for social cognition, there should include other factors that force infants' behaviors to be performed in social contexts. Since we suppose that the motor development should drive and enhance the higher cognitive development in infancy, the basic functions to detect spatiotemporal changes in their environments and others' actions might precede in the early stages of human cognitive development. This may contribute to the adjustment of one's own movements, and then to the differentiation into a higher cognition of reflecting internal mechanisms such as inferring others'

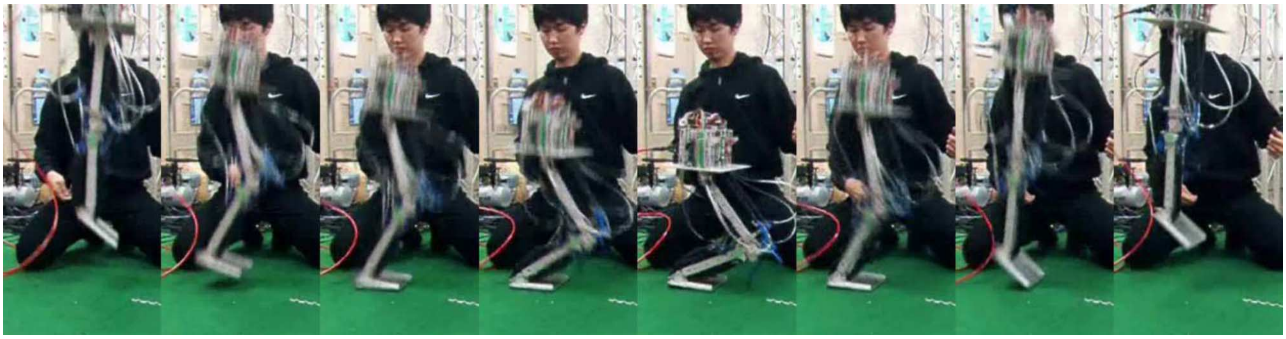


Fig. 5. A bouncing sequence. The robot can bounce up to five steps, which is quite stabilized even without any sensory feedback (from [41]).

intentions and observing any changes in the environment. Thus we expect such developmental features to be observable in early infancy as various kinds of interpersonal plays and physical interactions with peers, which might underlie the development of cortical system in ACC for coordinating rhythmic and discrete movements.

The above argument partially implies the developmental course from "reflex" to "higher order cognition" indicated by a blue arrow at the center in Fig. 3. The computational model corresponding to the above process is now under the construction in conjunction with observation study at our group.[3]

B. Musculoskeletal System as Physical Embodiment

These neural architectures for motor control are supposed to be not innate but learned through the physical interaction with the environment. However, the synthetic approach to study motor skill development has been facing difficulty due to the lack of appropriate physical mechanism since the conventional motor control using electromagnetic motors is limited to realize dynamic (high speed) motions that cause delay for control due to the limits of control frequency, as mentioned in Section II-A. Similarly, biological systems have some delay for neural information processing due to the limit of the velocity of neurotransmission. However, the musculoskeletal system works as no-delay system with nonlinear property, and consequently realize stable motions [49], [50]. Therefore, we should pay attention to such a mechanism, and McKibben pneumatic actuators have been receiving increased attention as biomimetic artificial muscles to generate dynamic motions with compliance like natural muscles of animals.

The human musculoskeletal system has a complicated structure consisting of bones, joints, ligaments, and muscles. A synergistic movement of such body parts emerges through interaction between such a complex body, the controller, and the environment. Not only the structure but also the physical properties of each component plays an important role to realize human's dynamic locomotion.

Blickhan's simple spring/mass model is one of the milestones for studying dynamic locomotion of animals [51]. As the model was so simple and easy to describe, storing and releasing energy during running, many researchers adopted it to analyze running (e.g., [52]–[54]). Based on a similar dynamic model, Raibert carried out pioneering research and developed a biped robot that had a springy prismatic joint that could run, jump, and perform somersaults [55]. If such a joint is used for walking as well, however, its compliance needs to be changed since the compliance suitable for walking is supposed to be different from that for running [56]. If the compliance changes, the locomotion mode of the robot may also change [57].

Natural animals change their posture and muscle tone according to their locomotion mode. As a result, animals can not only run but also realize other locomotion modes. Hyon *et al.* developed a running robot imitating the structure of a hind leg of a dog [58]. Iida *et al.* developed a human-like robot with several springs as muscles and investigated emergence of walking and running [57]. Realized locomotion modes in their studies were, however, relatively limited since their robot had fixed springs. Hurst *et al.* developed a monopod with tunable springs to realize a wider range of dynamic locomotion [59]. Since they adopted huge fiberglass springs, the realized structure is relatively simple. Vanderborght *et al.* developed a biped robot "Lucy," driven by pneumatic artificial muscles, which have the possibility to change the joint compliance [60]. However, they did not utilize variable compliance to realize different locomotion modes.

Hosoda and Ueda [61] focused on external rotation of the hip joint as important morphological feature for infant development. They assumed that the external rotation plays an important role in the emergence of biped walking. They worked on this assumption by deriving kinematic equations and real experiments.

Narioka and Hosoda [62] built a whole body humanoid driven by pneumatic artificial muscles and realized biped walking, utilizing its dynamics without using traditional trajectory-based technique. To realize biped walking with such a robot, compliance in the ankle plays an important role. They proposed adopting rollover shape to determine the ankle compliance, which was supposed to be one adaptability measure for human walking [63]. As a result, they expect to understand the adaptability principle underlying both humans and robots based on their dynamics.

Niiyama and Kuniyoshi [64], Hosoda *et al.* [41], and Takuma *et al.* [65] developed bouncing robots to realize vivid, dynamic motions with very low computational cost. Fig. 5 shows bounce motion realized in the work described in [41]. They showed experimentally that the biarticular muscles strongly governed the

---

[3] http://www.jeap.org/.

TABLE II
LOCOMOTION WITH COMPLIANCE

key issue	study
a simple spring/mass model for dynamic locomotion:	Blickhan [52]
running quadruped robots:	Koditscheck et al. [53] Ahmadi & Buehler [54]
biped walking and running simulation by springy legs:	Seyfarth et al. [55], [57] Iida [58]
jumping and running robots by prismatic elastic joints:	Raibert [56]
a jumping robot with mechanical compliance:	Hurst et al. [60]
biological-inspired structure:	
- walking humanoid:	Vanderborght et al. [61] Narioka & Hosoda [63] Hosoda & Narioka [64]
- jumping with bi-articular muscles:	Hyon & Mita [59] Hosoda et al. [42] Niiyama & Kuniyoshi [65]
- multi-modal locomotion:	Takuma et al., [66]
developmental aspect of walking:	Hosoda & Ueda [62]

coordinated movement of its body, and therefore a simple controller could realize stable bouncing. These robots indicate that control and body structure are strongly connected, that is, we can interpret that the body itself has a role of calculation for the body control [66]. One extreme and typical example is passive dynamic walkers that realize walking on a slope without any explicit control or actuation [67]. This is important from a viewpoint of energy consumption (resource bounded or fatigue).

Table II gives a summary of key issues related to locomotion with compliance.

C. CB^2 as a New Research Platform for CDR

Another type of pneumatic actuator is air cylinder type that is used for a research platform, CB^2, a child robot with biomimetic body for CDR [68]. CB^2 was designed, especially to establish and maintain a long-term social interaction between human and robot. The most significant features of CB^2 are a whole-body soft skin (silicon surface with many tactile sensors underneath) and flexible joints (51 pneumatic actuators). Fig. 6 shows CB^2 and its skeleton structure.

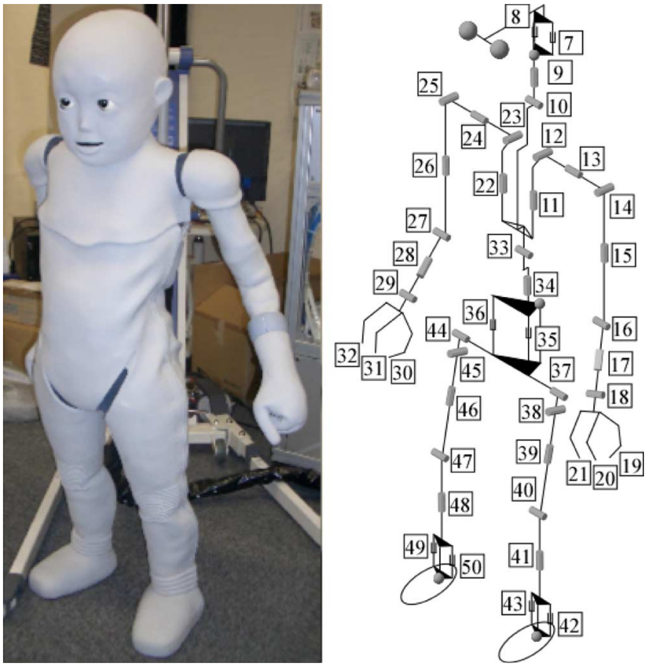


Fig. 6. CB^2: a child robot with biomimetic body as a research platform for CDR.

Table III compares CB^2 with the typical humanoid robots for studying human–robot communication and human development with respect to the presence of several features which are required for natural and tight human–robot interaction. The joint flexibility and the soft, sensitive skin provide tight interaction with humans. The human-like motion owing to the actuators mounted throughout whole-body and the child-like appearance invite child-directed behaviors from humans.

Ikemoto *et al.* [81] applied a new control system taking advantage of inherent joint flexibility to a case study of physical, direct interaction between a human and a robot where CB^2 attempts to rise up with human assistance. The conventional trajectory-based motion control methods are not able to be applied since precise position control for 51 pneumatic actuators is almost impossible. Instead, a simple control system with three postures as initial, intermediate, and final ones without caring about the precise position control is able to allow CB^2 to rise owing to its physical body structure's being similar to a human's, such that it absorbs the position errors to some extent using its joint flexibility.

The top of Fig. 7 shows an image sequence of a successful case of rising. A new measure was invented to clarify the difference between successful and unsuccessful cases based on an idea that at the beginning of the trial, robot motion may have some delay since a human will start to move earlier, but their motions should be synchronized toward the final posture. Then, the temporal correlation between joint velocities of CB^2 and change of joint positions of humans obtained by motion captures were measured. The bottom of Fig. 7 shows cross-correlation in terms of time lag (horizontal axis) during the time course of the trial (vertical axis). At the beginning (bottom line), the robot motions have a time lag (about 0.4 s) from the onset of the human motion, and gradually catch up toward the final posture (high correlation regions are linked). In unsuccessful cases, these regions are not connected. This measure might be useful when evaluating the physical interaction between a human and a robot when they cooperate to accomplish the given task.

TABLE III
COMPARISON OF HARDWARE SPECIFICATION

	Joint flexibility	Soft and sensitive skin	Actuators mounted throughout whole-body	Human like appearance	Child size
Robovie-II [70]	No (Electrical motors)	No	No (Upper body)	No	No
Infanoid [71] BARTHOC [72]	No (Electrical motors)	No	No (Upper body)	No	**Yes**
iCub [73]	No (Electrical motors)	No	**Yes**	No	**Yes**
Robovie-IIF [74]	No (Electrical motors)	**Yes**	No (Upper body)	No	No
ASIMO [75] QRIO [76]	No (Electrical motors)	No	**Yes**	No	**Yes**
SAYA [77]	**Yes** (Pneumatic actuators)	No (No tactile sensor)	No (Head)	**Yes**	No
CB [78]	**Yes** (Hydraulic actuators)	No	**Yes**	No	No
Repliee R1 [79]	No (Electrical motors)	No (No tactile sensor)	No (Head)	**Yes**	**Yes**
Repliee Q2 [80] Geminoid [81]	**Yes** (Pneumatic actuators)	**Yes**	No (Upper body)	**Yes**	No
CB^2 [69]	**Yes** (Pneumatic actuators)	**Yes**	**Yes**	**Yes**	**Yes**

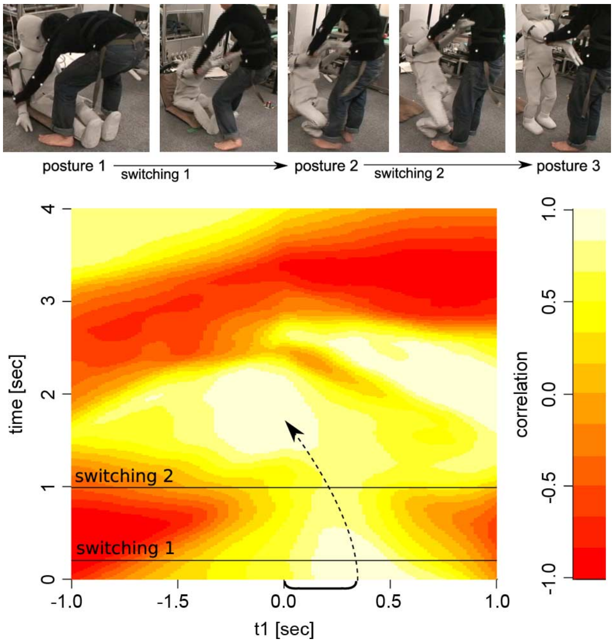


Fig. 7. Result of smooth interaction by an expert.

V. BODY REPRESENTATION, MOTOR REPRESENTATION, AND SPATIAL PERCEPTION

Body representations have been called the "body schema," an unconscious neural map in which multimodal sensory data are unified, and "body image," an explicit mental representation of the body and its functions [82]. Sometimes, it is called "motor image," which suggests a strong connection with motions. Ramachandran's famous book tells us how our brains are easily tricked by controlling the timing of motions such as synchronous rubbing of the noses to create the illusion of nose extension [83]. This implies that motions deeply participate in the developmental process of sense and perception.

In this section, conventional body representation in robotics is given, then CDR approach is introduced, and lastly neuroscientific approach is briefly given.

A. *Conventional Body Representation in Robotics*

Conventional body representation in robotics has mainly dealt with a morphology of the skeleton system, that is, link and joint structure with their parameters given. A more adaptive approach makes a robot estimate these parameters based on its experience in the environment [84]–[86]. The latter approach is closely related to modeling of the human body representation because recent brain and medical studies revealed that biological systems have a flexible body representation, so-called body image. Ramachandran showed that patients suffering from phantom limb pain could alleviate their pain by observing the visual feedback of the good limb in mirror box, and suggested that the cortical representation of the body might have been restructured [83]. Iriki *et al.* showed that the receptive field of the bimodal (somatosensory and visual) neurons in the intraparietal cortex is extended when monkeys use a tool to obtain food [87]. Moreover, these body images are thought to represent the relationship between an animal's own body and the external world. This may suggest that body image is the spatiotemporally integrated image of various modalities, such as auditory and visual perceptions and somatic (including tactile) sensations as well.

As mentioned before, neonatal imitation [28] has been a hot topic causing a controversial argument between "innate" and "learned." Meltzoff and Moore proposed the active intermodal mapping (AIM) model to explain this form of early imitation [89]. In their model, organ identification, through which newborns can associate the sensory perception of invisible parts with the features of parts of others in visual information, is a prerequisite. Breazeal *et al.* proposed a model of facial imitation based on the AIM model [90]. In this model, in order to acquire the organ identification ability, the robot learns the relationship between the tracking data of features of the face of the other robot and the joints of its own face when imitating another

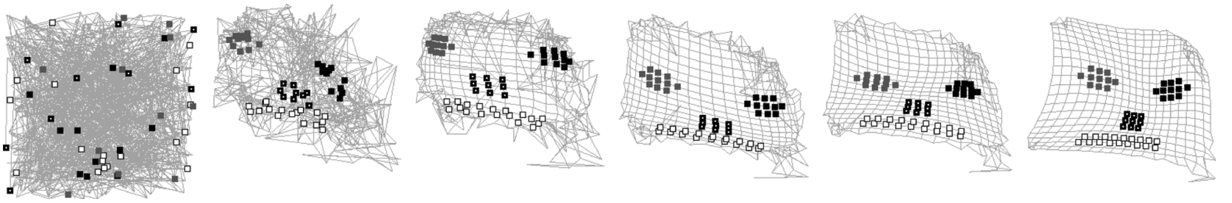


Fig. 8. The configuration of the tactile sensor units from the random state (leftmost: the first step) to the final one (rightmost: the 7200th step) where gray squares, black ones, empty squares with thin lines, and empty squares with thick lines correspond to right eye, left one, mouth, and nose (from [88]).

robot. However, it remains unclear as to how infants understand that their gestures are the same as those of the person being imitated.

B. Synthetic Approaches to Body Representation and Frame of Reference

Recent studies revealed the possibility of fetus learning in the womb through the emergence of fetal motility and sense as mentioned before. Thus, it does not seem unreasonable to suppose that infants acquire a primitive body image through experiences in the womb.

CDR approaches this issue from different direction. Nabeshima *et al.* [91] proposed a model to explain the behavior of the neuron observed in the experiment of Iriki *et al.* [87]. In their model, a robot detects the synchronization of the visuo-tactile sensations based on an associative memory module and acquires a body image. Yoshikawa *et al.* [92] proposed a model in which a robot develops an association among visual, tactile, and somatic sensations based on Hebbian learning while touching its own body with its hand. Fuke *et al.* [88] proposed a learning model that enables a robot to acquire a body image for parts of its body that are invisible to itself. The model associates spatial perception based on motor experience and motor image with perception based on the activations of touch sensors and tactile image, both of which are supported by visual information. Fig. 8 shows the configuration of the tactile sensor units from the random state to the final one where gray squares, black ones, empty squares with thin lines, and empty squares with thick lines correspond to right eye, left one, mouth, and nose, respectively.

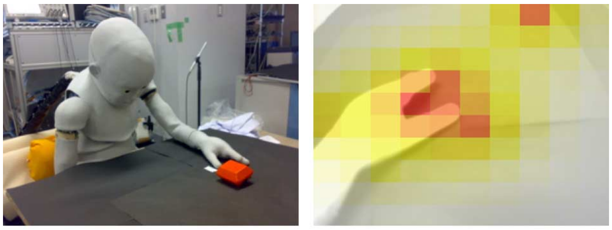
Finding self body from the sensory data that may include both the self and others (object or agent) is an issue for infants not only to represent their own body but also to learn to manipulate objects as tools. Asada *et al.* [3] proposed a method where a robot finds its own body in the visual image based on the change of sensation that correlates with the motor commands. Yoshikawa *et al.* [93] proposed a model in which a robot can detect its own body in the camera image based on the invariance in multiple sensory data. Furthermore, Stoytchev [94] proposed a model that enables a robot to detect its own body in a TV monitor based on the synchronization of the activation of vision and proprioception. Hersch *et al.* [95] proposed an algorithm through which a robot learns joint positions and orientations based on the information of the observing hand's positions represented in both the head-centered and the hand-centered reference frames. Hikita *et al.* [96] presents a method that constructs a cross-modal body representation from vision, touch, and proprioception. When the robot touches something, the activation of tactile sense triggers the construction process of the visual receptive field for body parts that can be found by visual attention based on a saliency map[4] and consequently regarded as the end effector. Simultaneously, proprioceptive information is associated with this visual receptive field to achieve the cross-modal body representation. The computer simulation and the real robot results are comparable to the activities of parietal neurons found in the Japanese macaques [98]. Fig. 9(b) shows the acquired visual receptive fields with [(c) and (d)] and without a tool [(a) and (b)].

Most body representation in CDR approaches adopt sensor-based representation such as retinotopic coordinates rather than the head-centered coordinates. Therefore, the correct integration is not accomplished if the robot moves its head, which is also the case for infants. This problem will be solved if the robot acquires head-centered coordinates. The head-centered coordinates will be obtained by associating eyeball angles and visual information. In human brains, the VIP neurons found in the parietal lobe are supposed to code for the head-centered representation and also to connect visual and tactile sensations (face) through "hand regard" behavior [100]–[102].

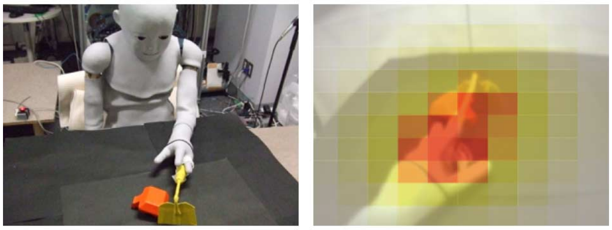
Fuke *et al.* [99] considered which information human infants might regard as reference information according to neurophysiological and cognitive findings. Then, they proposed a learning model in which a robot acquires not only the head-centered reference frame but also the cross-modal representation of the face based on raw sensory data, by focusing on the behavior that can be observed in the human developmental process. The acquired cross-modal representation corresponds to the actual properties of VIP neurons found in neuroscience.

Andersen [103] found the neurons in the monkey parietal cortex area, that is, lateral intraparietal (LIP) area, that combine three kinds of signal: the position of the stimulus on the retina, the positions of the eyes in the orbit, and the angles of the head. The LIP area connects to the VIP area [104] and is reported to have both eye-centered and head-centered visual receptive fields [105]. The head movement is not dealt with by Fuke *et al.* [99], but it can be assumed that the head-centered visual space corresponds to the LIP area as shown in Fig. 10, where correspondence between regions in the brain and spaces in the model by Fuke *et al.* [99] is shown.

[4] The saliency map is proposed based on biologically plausible architecture by Itti *et al.* [97]. The map is constructed by combining several kinds of features in the visual image.



(a) without a tool (b) receptive field of (a)



(c) with a tool (d) receptive field of (c)

Fig. 9. The acquired visual receptive fields (c) and (d) with and (a) and (b) without a tool (from [96]).

C. Neuroscientific Approach to Frame of Reference

In parallel with synthetic approaches to body representation and spatial perception, neuroscientific studies corresponding to the synthetic ones are conducted.

Ogawa and Inui [106] conducted a functional magnetic resonance imaging (fMRI) study of manual tracking. In this experiment, participants tracked a sinusoidally moving target with a mouse cursor. In some trials, vision of either the target (externally generated) or the cursor (self-generated) movement was transiently occluded, while subjects continued tracking by estimating the current position of either the invisible target or the cursor on screen. The results revealed lateralization of brain activity depending on whether the target or cursor became invisible: the right and left posterior parietal cortex (PPC) showed greater activation during occlusion of target and cursor movements, respectively (Fig. 11). This finding indicates that an object whose movement is congruent with our own body motion (cursor) is estimated predominantly in the left hemisphere of the brain, whereas an externally generated movement, whose motion is not related with our own body motion (target), is predicted mainly in the right hemisphere. Their previous study also indicates that visual error between internally estimated and actual visual feedback of our effector's movement is evaluated in the right intraparietal sulcus, and the error is properly integrated into internal estimation in the right temporo-parietal junction [107]. Their results also indicate that the presupplementary motor area is related to visuo-motor imagery irrespective of whether the occluded motion is self- or externally generated [106].

On the other hand, one of the developmental goals during infancy is to establish various frames of reference to apprehend certain objects in the external world and to adopt an appropriate frame of reference depending on the situation to adapt to the environment. There are two types of frame of reference, egocentric and allocentric, both of which are deeply concerned with body schema. We can easily use these frames of reference to realize various higher cognitive functions [108]. During development, it is most important to establish the two frames of reference and to learn mutual transformation between them. Based on a number of neuroimaging studies, Inui [108] proposed a hypothesis that the left parietal lobe is assigned to the egocentric description, while the right parietal lobe is assigned to the allocentric description. In addition, he suggests that the left parietal lobe functions to project an external stimulus to self-body, while the right parietal lobe functions to project self-body to an external object or other's body. He also presumes that body images are generated with a physical prediction control system.

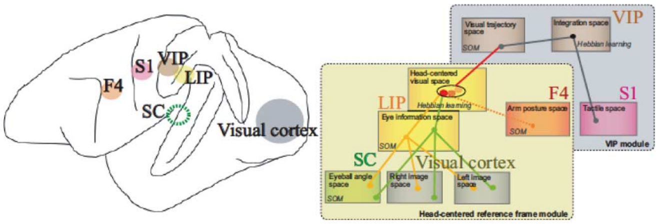


Fig. 10. Correspondence between brain regions and representation spaces proposed by [99].

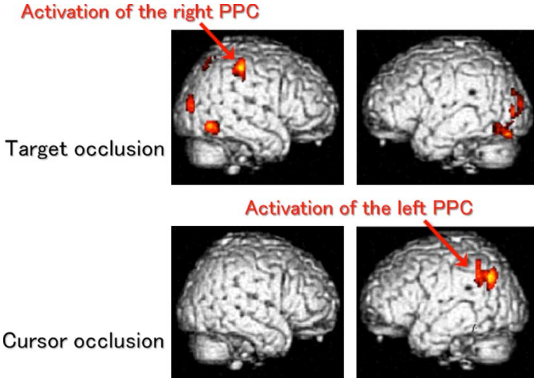


Fig. 11. The activity in the right PPC was observed in target occlusion condition without the left PPC activity and vice versa for cursor occlusion condition.

VI. FROM EMERGENCE OF SOCIAL BEHAVIOR THROUGH INTERACTIONS WITH CAREGIVER TO DEVELOPMENT OF COMMUNICATION

In the previous sections, we have reviewed the developmental processes for individuals and the relationship between objects and individuals and have shown the correspondences to the brain regions in terms of functions as much as possible. As mentioned before, the medial frontal cortex is closely related to mind development and social cognition [47]. However, it seems that a more global network of the brain works together for such development and cognition, and, more importantly, interaction triggered by caregivers as one of the environmental factors plays an essential role in the developmental process for communication. Here, we deal with the issues of early communication, vowel imitation, joint attention, and empathy development.

A. Early Communication

Knowledge about human adaptability in communication is not sufficient when applying the theories to the design of communicative robots. It will be helpful to make a model and validate it using a virtual infant robot so as to investigate the underlying mechanism about how humans develop adaptive communication ability [109].

After birth, the intimate communication between the infant and the caregiver starts. At first, the communication seems very reflexive. However, as the development proceeds, the infant seems to learn to understand how to respond to the caregiver's behavior.

Infant studies show that infants become sensitive to the regular behaviors of caregivers by four months after birth [110]. In the same month of development, infants begin to adjust the timing of the communication with their own mothers and, furthermore, generalize the timing so that they show the same response to unknown persons whose timing is the same as their mother's [111]. When the interchange begins between infants and caregivers, the infants develop an ability to predict interactions with social partners [111].

Rochat *et al.* [110] investigated the responses of two-, four-, and six-month-old infants to regular and irregular peekaboo communication. Two-month-old infants showed equal attention and smile levels both to regular and irregular peekaboo communications. However, 4 and 6 month old infants showed less attention and more smiles to regular peekaboo than to irregular peekaboo.

These experiments indicate that after four months, infants 1) memorize the behavior of caregivers and 2) adjust the timing with expectation of the next behavior based on the memory. Ogino *et al.* [112] hypothesize that two emotional processes take important roles. The first is the memorizing. In neuroscience, it is observed that the emotional stimulus of the amygdala affects the pathway from the cortical to hippocampus, and the memory of the events before and after the stimulus is strengthened [113], [114]. The second process is the prediction of the reward. In neuroscience, it is well known that the dopamine neuron in basal ganglia takes an important role in reward prediction [115].

In developmental psychology, peekaboo is treated as one of the communication styles in which infants adjust to the emotions and affections of caregivers, and often used in experiments to examine the infant's abilities to detect regular social behavior in communications and to predict certain behaviors [110].

Mirza *et al.* propose an interesting model in which the peekaboo game emerges as a robot's behavior based on its own experience and the stimulus from the environment [116]. This model might explain one of the contributing factors of how infants begin to play the peekaboo game of their own volition. Ogino *et al.* [112] proposed a communication model for a robot to acquire peekaboo based on the reward prediction. The system keeps the sensor data in short-term memory and transfers them to long-term memory when the value of internal state corresponding to the emotion is increased. Once the memory is formulated, the sensor data are compared with it and the robot expects the regular response of the caregiver.

In these synthetic approaches to early communication of infants, it is still an interesting question of how the turn-taking game emerges based not only on the self-experience but on the recognition of others. It is only after the sixth month that infants begin to play the peekaboo game. Between around six months to one year after birth, infants acquire the knowledge that objects continue to exist even when they become invisible. This concept is called "object permanence" [117] and is related to an infant's faculty for image generation or motor prediction in the brain. These abilities of prediction are also considered to develop along with the infant's acquisition of goal-directed movements (reaching or reaching to grasp). In addition, predictive control mechanisms may play important roles in the development of nonverbal communication such as pointing or imitation.

It is also in the same period that the shared attention and the imitation of behaviors begin. In that case, a more interesting question is how the emotion model of others affects to the acquisition of communication. We touch upon this in later sections.

B. *Action Execution and Understanding*

Infants start to reach and grasp, or to imitate actions involving objects between about six to nine months after birth. Whereas infants under 12 months pay more attention to the movement of an action itself, those over 12 months can imitate the action regarding its goal or effect [118]. It is also indicated that even a three-month-old child can show understanding of an action regarding its goal after his/her own experience of object retrieval [119]. These findings indicate that infants understand others' actions based on their own action experience during development. Previous studies in neuroscience have also revealed that a specific brain region for action execution is recruited for the understanding of others' actions, which is called mirror neuron system (MNS) [120]. However, the relationship between MNS and different levels of action understanding remains unclear.

Ogawa and Inui [121] investigated whether or not there is a common neural coding of action execution and recognition dependent on different action levels. In this experiment, two movies, showing an action of grasping a pen and putting it into one of two cups located on either side of a table, were presented. Participants judged whether the first action was matched with the second action, regarding the following four aspects: 1) CUP: the cup in which the pen was placed (left or right cup); 2) GRASP: how the pen was grasped (thumb pointing upwards or downwards); 3) HAND: the hand used to grasp the pen (left or right hand); and 4) PATH: the rotating path of hand movement (clockwise or counterclockwise). Results showed that different brain regions were activated under each condition: medial prefrontal cortex [47] in CUP; left anterior intraparietal sulcus [122] in GRASP; bilateral superior parietal lobes [123] in HAND; and left premotor and primary motor areas [124] in PATH. The current study indicates that distinct brain regions are involved in observation of different aspects of transitive actions, consistent with a hierarchically organized visuo-motor network of the observer's own actions.

C. *Development of Vocal Imitation*

It has been reported that children of eight months can imitate an adult's single vowel [125]. To reach such a developmental milestone of imitation, infants should not only acquire sensorimotor mapping to vocalize sound but also find the correspondence of utterances between themselves and their caregivers. It has been a central interest of developmental science how infants acquire the abilities underlying these requirements. Infants' ability of listening to adult voices appears in a language-independent manner from birth and gradually adapts to their mother tongue [126]. Meanwhile, infants' utterances are first quasi-vocalic sounds that resemble vowels and gradually adapted to their caregiver's ones [127] along with descent of the epiglottis [128]. Therefore, it seems likely that vocal interaction with their caregivers is needed for such an infant to adapt its vocal system to the caregivers' language. However, how and what kinds of interaction among infant's learning mechanisms and the caregiver's behavior are essential for the processes is still unclear.

On the other hand, recent imaging technology has started locating early sensitivities for language input in the infant brain [129], [130]. However, it remains difficult to investigate the links among these sensitivities and caregiver's interaction through developmental course because of the limitations of current imaging technology. Since similar difficulties also exist in other approaches based on observation due to ethical problems to control infant development, synthetic approaches are expected to contribute to find the missing links.

Some synthetic studies have been conducted to model what happens in the vocal babbling period that has often been considered to have important roles in speech acquisition. Guenther *et al.* [131] have developed a neural network model called DIVA model that learns sensorimotor mapping through random exploration resembling infant's babbling, and argued its correspondences to results of adult fMRI work. Westermann and Miranda [132] have developed another model that incorporates physical embodiment in sensorimotor learning so that it can acquire reliable vowel articulations through a babbling period while being exposed to ambient language. Kanda *et al.* [133] have proposed a recurrent neural network model that automatically segments continuous sounds of vowel sequence to learn sensorimotor mapping and pointed out the importance of incorporating self articulation in the segmentation process. Hörnstein and Santos-Victor [134] have considered multiple phases through babbling period by regarding the caregiver's behavior during that time and argued the same importance on recognition of others' vowels. However, the above synthetic studies have not paid much attention to the caregiver's behavior, which has started from birth and seems essential in vocal development.

Kokkinaki and Kugiumutzakis have reported an important characteristic of caregiver's behavior for infants to learn such correspondences: parents imitate their infants at a high rate in the first six months [135]. As implied from other observations where imitation of an infant's utterances by caregivers is induced by the infant's vowel-like utterances [136] and inversely encourages such utterances [137], such parental imitation or being imitated might play an important role in the developmental process of vocal imitation. The importance of being imitated has been demonstrated in synthetic studies of computer-simulated vocal agents, although they did not directly aim at modeling infants' development. It has been shown that a population of learning agents with a vocal tract and cochlea can self-organize shared vowels among the population through mutual imitation [138], [139]. However, these previous works have assumed that the infant model could produce the same sound as caregivers if it learned proper parameters of articulation unlike the situation of infants due to their immaturity. In other words, they have paid less attention to another developmental hurdle of finding the correspondence of utterances between themselves and their caregivers.

On the other hand, Yoshikawa *et al.* [140] have addressed this issue in human–robot vocal interaction and demonstrated the importance of being imitated by the human caregiver, whose body is different from the robot's, as well as subjective criteria of the robot such as ease of articulation. With a similar experimental setting, Miura *et al.* [141] have argued that being imitated by a caregiver has two meanings: not only informing of its correspondence to the caregiver but also guiding the robot to performing it in a more similar way to how the caregiver performs.

Inspired by the previous work [141], Ishihara *et al.* [142] have computationally modeled an imitation mechanism as a Gaussian mixture network (GMN) parts of which parameters are used to represent caregiver's sensorimotor biases such as perceptual magnet effect [143], which is adopted in the computer simulation of Oudeyer to evolve a population to share vowels [144]. Perceptual magnet effect indicates a psychological phenomenon where a person recognizes stimulus as more typical of closer categories that the person possesses in his or her mind. They have conducted a computer simulation where an infant and the caregiver imitate each other with their own GMNs, where one for the infant is learnable and the other for the caregiver involves a certain level of magnet effects. They have found that caregiver's imitation with magnet effects could guide infants' vowel categories toward corresponding ones. Interestingly, the effectiveness of guidance was enhanced if there was what they call automirroring bias in the caregiver's perception so that she perceives the infant's voice as closer to the one that resembled her precedent utterance.

Fig. 12 shows the vocal imitation mechanism proposed by them [142], where automirroring bias and sensorimotor magnet effect are involved as two modules. Fig. 13(a) shows the learning result with two well-balanced biases, where blue and red dots indicate the caregiver and the infant voices, respectively, apexes of red pentagons represent target vowels of the infant, in other words, clearest vowels in her vowel region, and black dots represent infant vowel prototypes after learning. In (a), the infant vowels are almost correctly converged, while (b)–(d) do not seem to be. In (b), the sensorimotor magnet is missing, therefore not convergence but rather divergence of infant vowels and the caregiver's imitated voices is observed. On the other hand, in (c), automirroring biases is missing, and as a result, there is fast convergence but three of five vowels converged onto wrong locations. In the case of no biases, nothing happened, as shown in (d).

Although previous synthetic studies focusing on finding correspondence through being imitated have assumed that the infant is almost always or always imitated by the caregiver for simplicity, it is apparently unrealistic. In such more realistic situations, infants should become able to realize that they are being imitated. Miura *et al.* [145] have addressed this issue by considering a lower rate of being imitated in computer simulation and proposed a method called autoregulation, that is, active selection of action and data with underdeveloped classifiers of caregiver's imitation of infant's utterances.

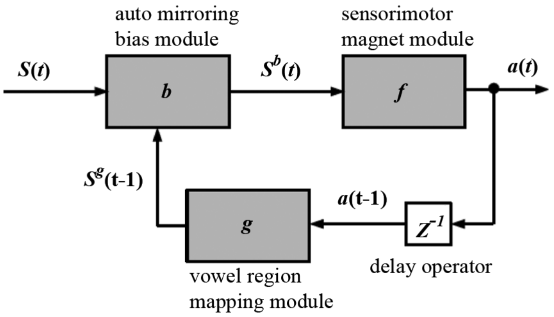


Fig. 12. Vocal imitation mechanism (from [142]).

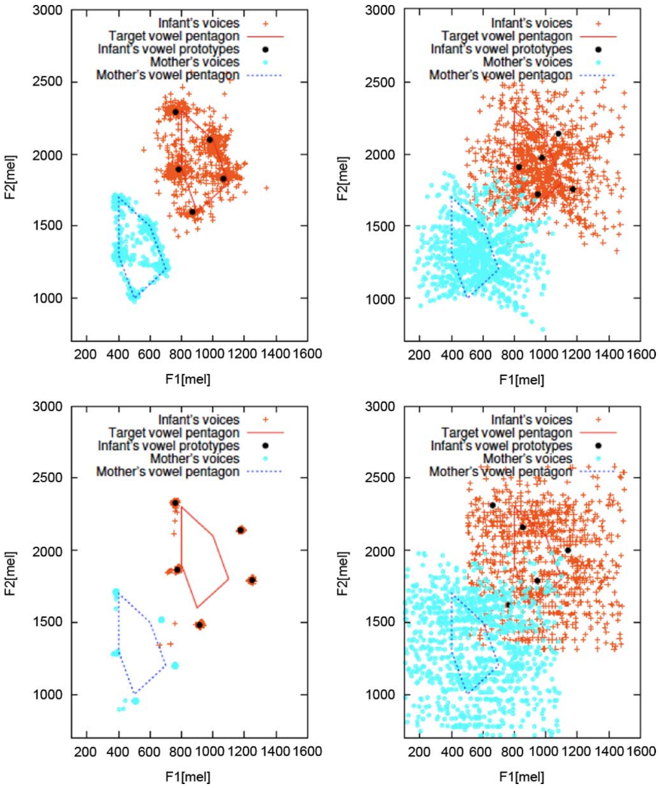


Fig. 13. Difference of learning results under several conditions. Apexes of red pentagons represent target vowels of infant, in other words, clearest vowels in her vowel region, and black dots represent infant vowel prototypes after learning (from [142]).

In many previous synthetic studies, the problems of learning a sensorimotor map and finding correspondence have always been coped with separately and under the situation where only vocal exchanges are assumed. However, to more faithfully model the developmental process, we should consider both problems under a more realistic situation of vocal interaction such as sharing attention or naming game.

D. Development of Joint Attention

Joint attention, that is, looking at the same object at which another is looking, is one of the bases for higher interpersonal cognition [146] and communication and has therefore been a central topic in developmental psychology [147].

In the following, we first mention the face preference of infants mostly supposed innate and argue any possibility of how it is acquired through learning based on the attempt to model the autistic children who are not good at communicating with people. Then, we review the normal developmental process of joint attention.

Infants are known to have a preference for the face-like patterns just after birth, and this ability is believed to be innate [148], [149]. This ability matures as they grow up, so that they can learn to distinguish and categorize different people in terms of various points [150]. However, recently, Fasel *et al.* mentioned that only 6 min are enough to collect the data to train the detector for face-like image, and it is still questioned whether or not facial preference is innate [151]. On the other hand, some people with autism spectrum disorders do not show any preference toward human faces. It is observed that some people with autism spectrum disorders have a weak preference to the eyes of others [152]. This distinctive difference of the preference of the attention from usual people is suspected as one of the causes why children with autism spectrum disorders fail to acquire communication skills. Actually, some therapy emphasizes the training of autism children to look at the other's face. Applied behavior analysis (ABA) therapy was first developed by Lovaas, and it is reported that social skills of children with autism spectrum disorders are improved to some extent through ABA therapy [153]. Although many techniques are proposed in ABA therapy, the basic idea is to classify social behaviors into the behavior elements and reinforce each behavior element by the reward.

It is reported that the attention of some children with autism spectrum disorders is less affected by motion information [154], and it is thought that this might be the cause of the failure of these children to acquire communication skills. Ogino *et al.* [155] modeled the process of ABA therapy of autistic children for eye contact as the learning of the categorization and preference through the interaction with a caregiver. The proposed model consists of a learning module and a visual attention module. The learning module learns the visual features of higher order local autocorrelation [156] that are important to discriminate the visual image before and after the reward is given. The visual attention module determines the attention point by a bottom-up process based on a saliency map and a top-down process based on the learned visual feature. The experiment with a virtual robot shows that it successfully learns visual features corresponding to the face first and then the eyes through the interaction with a caregiver. After the learning, the robot can attend to the caregiver's face and eyes as autistic children do in the actual ABA therapy.

The developmental processes of joint attention have been revealed by focusing on related social actions such as gaze alternation, i.e., successive looking between a caregiver and an object, social referencing, and pointing [157]. Imaging studies have started to reveal the developmental processes of neural substrates for joint attention. For example, early sensitivities for directing gaze toward the infant has been located in infant's medial occipital event-related potential [158]. It has been reported that electroencephalograph activities are predictive of latter development of joint attention: left parietal activities for responding to joint attention and left frontal as well as left and right central activities for initiating joint attention [159].

Due to ethical problems, however, it is not easy to proceed with longitudinal observation nor controlled experimentation of infant development. Therefore, it has been and is still a formidable issue to understand developmental processes such as what kind of learning architecture underlies it and how the caregiver's behavior affects and is affected by its progress. There have been a number of synthetic studies to reveal such missing links, that is, understanding the developmental processes of and/or through joint attention [160]–[164] (see also a thorough survey on this topic [165]). Some of them have focused on the issues of understanding how a learning agent can acquire a sensorimotor map to follow another's gaze from the interaction of a caregiver, which is a basic skill of joint attention. Nagai *et al.* [162] have argued how the learning process of joint attention is affected by the developmental changes in other aspects such as gradual maturation of visual perception, that is, how sharply it can observe the caregiver's face, as well as the changes in caregiver's instruction, that is, how tolerantly the caregiver rewards the robot performance.

Unlike the assumption in the previous work, it is unlikely that parents always pay attention to and reward their infants. Instead of relying on such rewarding, the role of the contingency inherent in the caregiver's looking behavior has recently become the focus of research [163], [164]. It causes the statistical tendency of finding something salient in the direction of the caregiver's gaze and can be utilized for learning the sensorimotor map of gaze following. Based on such contingency learning, Nagai *et al.* [163] have modelled the gradual extension of the gaze-followable area of infants from 12 to 18 months of age[166]. Triesh *et al.* [164] have modeled healthy, autistic, and Williams infants by changing characteristics of their preferences and simulated how development/nondevelopment of joint attention occurs in relation to the caregiver's behavior, namely, how the caregiver shifts gaze.

These previous works have focused on the skill of gaze-following instead of arguing how infants realize that such a specific skill should be learned. Sumioka *et al.* [167] have proposed an open-ended learning model of social actions by which an artificial infant reproduced the experienced contingency. To find which contingency should be reproduced, an information theoretic measure of contingency between two variables [168] has been extended to measure contingency inherent among an action and both prior and posterior sensations to the action. It has demonstrated that the proposed mechanism led to the seamless acquisition of social actions, that is, from gaze following to gaze alternation. Based on their fMRI study, Blakemore *et al.* [169] showed that the detection of intentional contingency between shapes whose movement was animate activated superior parietal networks bilaterally. These activations were unaffected by attention to contingency. Additional regions, the right middle frontal gyrus and left superior temporal sulcus, became activated by the animate-contingent stimuli when subjects specifically attended to the contingent nature of the stimuli.

Instead of visual cues, the caregiver's utterances about the focus of attention can be cues to perform joint attention for infants. However, the coordination among these different modalities does not seem to have matured in the early period of development: infants cannot refer to the gaze direction of adults when they learn word labels of objects from the adults' utterances until about 18 months of age [166], although they have already started acquiring word labels at this age [170]. The statistical mapping approach has also been adopted for modeling the development of word-to-object mapping [171], [172], which could be utilized for word-driven joint attention. However, these studies have focused only on either modality. On the other hand, gaze-driven joint attention has been shown to be necessary for statistical learning of word-to-object mapping [173] as infants older than 18 months of age do. Yoshikawa *et al.* [174] have addressed the issue of understanding how multimodal skills of joint attention can be interactively developed. They have proposed a method of simultaneous contingency learning not only of gaze-following mapping but also of word-to-object mapping and demonstrated that learning processes of these two social functions can facilitate the development of each other.

E. Empathy Development

Empathy is indispensable for communication. Although it is unclear how sympathetic feelings are evoked, facial expression is an important cue for eliciting empathy. Of the communication channels used by human beings, 55% are related to the face, 38% to the tone of voice, and 7% to verbal content [175]. However, it has not been clear how the capacity for empathy based on facial expression is formed. Infants instinctively respond to faces, and babies a few days old distinguish their mother's face from those of others after being contact with their parent for 11 or 12 h [176]. Conversely, an investigation of newborn facial expressions showed that the basic expressions are innate [177]. To realize natural communication between robots and their users, the processes underlying how these essential abilities are combined to elicit empathy must be clarified.

Human-like robots able to show distinct facial expressions have been developed [178], [179], but the facial expressions to be used in specific situations are specified explicitly by the designer in advance, leaving robots unable to adapt to nonspecified situations and unable to modify their internal state in response to the facial expressions of users. Breazeal *et al.* proposed a developmental model that enables a robot to derive the relationship between motor commands for its facial expressions and those of the caregiver's by imitating facial expressions during the robot's motor babbling [90]. Empathy, however, does not involve simply mimicking the facial expressions of others. More important is the ability to evoke the same internal state as others based on their facial expressions and vocal characteristics. Kobayashi *et al.* [179] proposed learning in which the robot categorizes a user's facial expressions under given emotional labels. This enables a robot to evoke the same emotional label felt when the caregiver touched the robot before. Again, however, emotional labels are fixed and the caregiver's active synchronization is not considered.

How do human children develop empathy through interactions with their caregivers? In developmental psychology, the caregiver behavior called "intuitive parenting" [180] serves as a "maternal scaffolding" upon which children develop empathy as they grow. A typical example is when caregivers mimic or exaggerate a child's emotional expressions [181]. This is considered a good opportunity for teaching children how to feel in real time [111], and most adults possess this skill. Children are thus able to understand the meaning of facial expressions and develop empathy toward others as the process is reinforced through emphasis on the facial expressions of their caregivers. This is because children empirically learn the connection between their internal state and the facial expressions of others.

Watanabe *et al.* [182] proposed a communication model that enables a robot to associate facial expressions with internal states through intuitive parenting by users who mimic or exaggerate a robot's facial expression. The robot strengthens the connection between its internal state and the facial expression associated with a particular state. The internal state of a robot and its facial expressions change dynamically depending on the external stimuli. After learning, facial expressions and internal states are classified and made to mutually correspond by strengthened connections.

What part of the brain is responsible for empathy? Using fMRI, Singer *et al.* [183] showed brain activity associated with understanding by the subject, of another person's pain. In this experiment, brain activity was observed when a subject was administered with a painful stimulus through a electrode on the back of the hand and when the subject was administered the stimulus associated with a simultaneous view of the subject's loved one. The ACC and the cerebellum are activated in both cases, in addition to somatosensory cortex, but the sensation of pain associated with observing the pain of someone else differs from that of experiencing pain oneself in the same region. These results suggest that the area associated with feeling the pain of others and oneself are the ACC and/or the cerebellum, and that human beings, although able to identify with pain felt by others, experience the two differently.

F. Toward Verbal Communication

Human infants learn new words at an incredible rate from around 18 months, and they acquire a vocabulary of 1000 to 2000 words by the time they are two years old [184]. This is called "language explosion" or "lexical explosion" and is one of the biggest mysteries of human cognitive developmental process.

The existing bottom-up approach in machine learning to lexicon acquisition has focused on the symbol grounding problem, in which the problem treated is how to connect sound information from a caregiver and sensor information that a robot captures from the environment [185]–[188]. A typical method proposed in these studies is based on the estimation of the co-occurrence probabilities between the words uttered by a caregiver and the visual features that a robot observes. In these experiments, training data sets are given by the caregiver, and the robot passively learns them.

However, such a statistical method does not seem sufficient to explain the lexical explosion. It is observed that human infants can acquire the lexical relationship between the meaning and the uttered word from only one teaching, even though there are many other possibilities. Cognitive psychologists have proposed that infants utilize some rules or constraints to acquire lexicon efficiently. Markman [189] proposed the "whole object" constraint and the mutual exclusivity constraint. Landau *et al.* [190] proposed the "geometrical" constraint. The word order can be used for constraining the meaning of the words, and some methods are proposed that use grammatical information to acquire the lexical relationship and to categorize the acquired words [191], [192].

Moreover, infants are not passive creatures. They actively and intentionally interact with the environment around them [111]. The period in which an infant starts to learn language overlaps with the onset of walking. The existing methods proposed in machine learning have neglected this active attitude of infants, and training data are passively received by the infants. It is well known that infants have selectivity for novel things and events. It is shown from many observations that they look longer at novel things than at known ones. This selectivity is thought to take an effective role in acquiring information for new events and so in language acquisition.

The active selection of motions including visual attention might play an important role in lexicon acquisition. It is important to make a curiosity model with which an agent decides how to react to the environment depending on its current knowledge so that it can acquire necessary information. Saliency is one of the fundamental factors for making this conscious and subconscious motivational process. Saliency is supposed to be evaluated by comparison with something in novelty and frequency. Walther *et al.* [193] proposed a visual attention model in which saliency level is calculated based on the spatial comparison with surrounding features.

Ogino *et al.* [194] focused on the temporal aspect of saliency, which is evaluated based on temporal comparison in the short-term and long-term memory of an agent, and proposed a lexical acquisition model in which saliency, evaluated based on a robot's experience, affects the visual attention and learning rate of a robot. A robot evaluates saliency for each visual feature of observed objects depending on habituation and learning experience. The curiosity based on the evaluated saliency affects the selection of objects to be attended and changes the learning rate for lexical acquisition.

A central issue in cognitive neuroscience concerns how distributed neural networks in the brain that are used in language learning and processing can be involved in non-linguistic cognitive sequence learning. Recently, Dominey *et al.* proposed a neural network model in which several areas in the prefrontal cortex dynamically interact with each other [195], [196]. This model can explain many data that have been found in neurophysiological [197]–[200], neuropsychological [201], and psychological fields [202], including artificial grammar learning [203] and conceptual grounding [204]. The main part of the model is the known cortico-striato-thalamo-cortical (CSTC) neuroanatomy of the human language system. It is assumed in the model that structural cues encoded in a recurrent cortical network in BA47 (BA stands for Brodmann area) activate a CSTC circuit to modulate the flow of lexical semantic information from BA45 to an integrated representation of meaning at the sentence level in BA44/6 [196].

VII. DISCUSSION

We have given an overview of the various aspects of cognitive development, proposed the idea of the developmental model, and introduced various kinds of experiments and applications, as briefly shown in Fig. 3. Real robot implementations,

TABLE IV
KEY ISSUES TO EMERGE SOCIAL BEHAVIORS

social behavior	neuroscientific evidence	psychological evidence	synthetic approach
early communication	memory strength:	timing:	Breazeal and Scassellati [110]
	Paz et al. [114], McGaugh [115]	Rochat et al. [111], Rochat [112]	Mirza [117]
	reward prediction: Schultz et al. [116]		Ogino et al. [113]
vocal imitation	descent of the epiglottis: Sasaki et al. [129]	8m infants imitate: Jones [126]	DIVA model: Guenther et al. [132]
	locating early sensitivities for language	adapt to mother tongue:	physical embodiment:
	input: Dehaene-Lambertz et al. [130]	Werker and Tees [127]	Westermann and Miranda [133]
	Gervain et al. [131]	perceptual magnet:	RNN: Kanda et al. [134]
		Kuhl and Meltzoff [128]	population: deBoer [139], Oudeyer [140]
		mutual imitation:	different body: Yoshikawa et al. [141]
		Kokkinaki and Kugiumutzakis [136]	unconscious guidance: Miura et al. [142]
		Masataka [137], Pelaez-Nogureas [138]	two biases: Ishihara et al. [143]
			detection of being imitated:
			Miura et al [146]
joint attention		interpersonal cognition:	
		Baron-Cohen [147]	
		Moore and Dunham [148]	
(face preference)	early sensitivity to face:	innateness: Johnson et al [149],	learnable: Fasel et al. [152]
	Farroni et al. [159]	Johnson [150]	
		categorization: Slater and Lewis [151]	
	neonatal preference:	weak preference of ASD:	less affected by the motion
	Farroni et al. [159]	Klin et al. [153]	information: Shic et al. [155]
		ABA therapy: Lovass [154]	ABA modeling: Ogino et al. [156]
(gaze alternation, social referencing, and pointing)		developmental processes:	Scassellati [161], Kozim et al. [162]
		Tomasello [158]	developmental: Nagai et al. [163]
	EEG activities: Mundy et al. [160]	gradual extension: Butterworth [167]	contingency: Nagai et al. [164]
		referencing: Baldwin [167]	ASD,WS: Triesh et al. [165]
		word labeling: Bates et al. [171]	survey: Kaplan and Hafner [166]
			open-ended learning model:
	Contingency in fMRI study:		Sumioka et al. [168]
	Blakemore et al. [170]		word-to-object mapping:
			Roy & Pentland [172], Yu et al. [173]
			gaze-driven joint attention:
			Yu et al. [174]
			multimodal joint attention:
			Yoshikawa et al.[175]
Empathy	fMRI study: Singer et al. [184]	face preference: Mehrabian [176]	distinct facial expressions:
	ACC and cerebellum	distinguish mother's face:	Matsui et al. [179]
		Johnson and Morton [177]	imitating facial expressions:
		innate facial expressions:	Breazeal et al. [91]
		Rosenstein and Oster [178]	categorize facial expression:
		intuitive parenting:	, Kobayashi et al. [180]
		Papousek and Papousek [181]	modeling intuitive parenting:
		caregivers mimic or exaggerate:	Watanabe et al. [183]
		Gergely and Watson [182]	
		good opportunity for teaching: Rochat [112]	
lexicon acquisition	Dapretto and Bookheimer [198]	lexical explosion: Pruett [185]	connect sound and sensory data
	Friederici et al [199]	whole object constraint and the mutual	Asoh et al. [186], Ishiguro et al. [187]
	Inui et al [200], Ogawa et al [201]	exclusivity constraint: Markman [190]	Steels and Kaplan [188], Iwahashi [189]
		geometrical constraint: Landau et al. [191]	grammatical information:
		active agent: Rochat [112]	Roy [192], Toyomura and Omori [193]
		novelty and frequency: Walther et al. [194]	temporal aspect of saliency:
		Caplan [202], Saffran et al. [203]	Ogino et al. [195]
		artificial grammar learning: Reber [204]	Dominey et al.[196], [197]
		conceptual grounding:	
		Hirsh-Pasek and Golinkoff [205]	

computer simulations, psychological experiments with robots or computer simulation, and brain imaging studies are shown as support for the model.

Although we attempted to cover the full range of research topics of cognitive development from fetal simulation to the beginning of communication, we might have missed a number of important issues to be dealt with. Including those issues, we review the whole process.

In the fetal simulation [29], introduced as a model of individual development, the processes and/or consequences of the interaction between neural-musculoskeletal model (brain and body) and the external environment are reflected in the brain development. This indicates that body and brain are not separable but instead tightly coupled and developed through the interaction with the external environment (in this case, the womb). In this sense, we say "body shapes brain" [9]. The current model is still very simple and missing many other brain regions, sensory organs, and the details of body parts. By adding these extra details, more realistic simulations can be done through mutual feedback with neuroscience, developmental psychology, and other related disciplines.

Another extension is to connect with real robot experiments and real infant studies. The research group of the JST ERATO Asada Synergistic Intelligence Project[5] developed prototypes for baby robots based on McKibben pneumatic actuators [205] and tactile sensor suits for a caregiver and a baby in order to measure the mother–infant physical interaction in holding [206]. Some preliminary results are given, but more improvements for the baby robots and a deeper analysis of the data captured in holding are expected.

With regards to the development of motor skills, we focus on the aspect of hardware such as actuators, tactile sensors, and whole body research platform CB^2 for CDR because we put emphasis on the physical embodiment, the central idea of the developmental pathway from motor skills to cognitive functions, and therefore we cannot skip the issue of such equipment for CDR to attack the main issue of cognitive development of humans and robots. McKibben pneumatic actuators and other air cylinder type actuators are found to be useful in generating dynamic and flexible motions compared to conventional electromagnetic motors, and to experimentally verify how a human-like musculoskeletal system works. However, the pathway from motor skills to cognitive functions has not been clear. Observation studies (e.g., [207]) imply the connection between motor experiences and cognitive development, but its underlying mechanism is still unclear. How does motor skill development relate to cognitive development, do they "trigger each other" or "interfere"? In addition to the hardware improvements, a new experimental scheme to model the pathway seems necessary.

Body/motor representation and spatial perception is one of the most fundamental issues of CDR, and imaging studies suggest the brain regions related to these representations and cognitive functions, but it is difficult to see from these studies how these functions develop in the brain. Although a number of synthetic approaches were shown to address this issue, each of them has its own assumptions and limitations that do not always match with the findings in neuroscience. More systematic efforts from both sides seem necessary to make the model hypothesized by synthetic approach more realistic and to set up imaging experiments so that the hypothesized model can be easily verified. "Object permanence" can be a good target to make such efforts since it has not been systematically attacked by synthetic approaches, although it is an important step to develop higher order cognitive functions.

Reaching and grasping are very important steps toward object manipulation and recognition, and therefore motor skill development and visual attention system should be well coordinated to realize such actions. In this paper, we have touched upon imaging studies for these actions, but not so much for synthetic approaches since we have been lacking good platforms suitable for developmental study, such as a finger–hand–arm system covered by soft skin with tactile sensors. In such a situation, Sandini's group has been doing developmental study for object recognition through grasping (e.g., [208]–[210]) with their hand–arm system. They started from motor and vision primitives and the system learned the sensorimotor mapping and consequently objects for so-called affordance. The improvement of the platform is necessary, which may lead to more analysis on the structural and functional correspondences between the modules in the system and the brain regions.

In the development of social behavior through the interaction between individuals, a caregiver as an active environmental factor explicitly and implicitly affects the cognitive development. Imitation is one of the most essential issues in cognitive development, and there have been many studies in different disciplines such as ethology, developmental psychology, neuroscience, and robotics (e.g., [211]–[213] and many more). Instead of a thorough survey of imitation in general, here we touched on neonatal imitation and others such as vocal imitation from a viewpoint of development. Table IV (on the previous page) shows a summary of key developmental aspects triggered by the caregiver to facilitate the emergence of social behavior. Regardless of each key aspect, the issue for infants is how to acquire the exact representation of "others," and this is expected to be obtained by elucidating the learning process of the mirror system.

The studies on developmental disorders such as Autism Spectrum Disorders (ASD) and Williams Syndrome (WS) seem useful to construct the computational model of cognitive development that is conversely expected to be able to explain the structure of such disorders. In this process, synthetic approaches such as CDR are very effective, and the meaning of such approaches become deeper, which will eventually lead to the creation of new scientific values of CDR. In conclusion, even though we still have many issues to attack, CDR seems the most promising approach to the design principle of cognitive development.

ACKNOWLEDGMENT

The authors would like to thank D. Thomas, a researcher with the JST ERATO Asada Synergistic Intelligence Project, for his helpful comments on the draft of this paper.

REFERENCES

[1] R. Brooks, "Intelligence without representation," *Artif. Intell.*, vol. 47, pp. 139–159, 1991.

[2] P. E. Agre, "Computational research on interaction and agency," *Artif. Intell.*, vol. 72, pp. 1–52, 1995.

[3] M. Asada, E. Uchibe, and K. Hosoda, "Cooperative behavior acquisition for mobile robots in dynamically changing real worlds via vision-based reinforcement learning and development," *Artif. Intell.*, vol. 110, pp. 275–292, 1999.

---

[5] http://www.jeap.org/

[4] R. Pfeifer and C. Scheier, *Understanding Intelligence*. Cambridge, MA: MIT Press, 1999.

[5] M. Asada, K. F. MacDorman, H. Ishiguro, and Y. Kuniyoshi, "Cognitive developmental robotics as a new paradigm for the design of humanoid robots," *Robot. Auton. Syst.*, vol. 37, pp. 185–193, 2001.

[6] R. Pfeifer and J. C. Bongard, *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge, MA: MIT Press, 2006.

[7] G. Sandini, G. Metta, and D. Vernon, "Robotcub: An open framework for research in embodied cognition," in *Proc. 4th IEEE/RAS Int. Conf. Human. Robots*, 2004, pp. 13–32.

[8] D. Vernon, G. Metta, and G. Sandini, "A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents," *IEEE Trans. Evol. Comput.*, vol. 11, pp. 151–180, 2007.

[9] Y. Kuniyoshi, Y. Yorozu, S. Suzuki, S. Sangawa, Y. Ohmura, K. Terada, and A. Nagakubo, "Emergence and development of embodied cognition: A constructivist approach using robots," *Progr. Brain Res.*, vol. 164, pp. 425–445, 2007.

[10] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen, "Autonomous mental development by robots and animals," *Science*, vol. 291, pp. 599–600, 2001.

[11] C. G. Atkeson, J. G. Hale, F. Pollick, M. Riley, S. Kotosaka, S. Schaal, T. Shibata, G. Tevatia, A. Ude, S. Vijayakumar, and M. Kawato, "Using humanoid robots to study human behavior," *IEEE Intell. Syst.*, vol. 15, pp. 46–56, Jul./Aug. 2000.

[12] , D. Purves, G. A. Augustine, D. Fitzpatrick, W. C. Hall, A.-S. LaMantia, J. O. McNamara, and L. E. White, Eds., *Neuroscience*, 4th ed. Sunderland, MA: Sinauer, 2008.

[13] J. I. P. de Vries, G. H. A. Visser, and H. F. R. Prechtl, "Fetal motility in the first half of pregnancy," *Clinics Develop. Med.*, vol. 94, pp. 46–64, 1984.

[14] J. L. Hopson, "Fetal psychology," *Psychol. Today*, vol. 31, no. 5, p. 44, Sep./Oct. 1998.

[15] S. Campbell, *Watch Me Grow, A Unique, 3-Dimensional Week-by-Week Look at Your Baby's Behavior and Development in the Womb*. London, U.K.: Carroll & Brown, 2004.

[16] H. Eswaran, J. Wilson, H. Preissl, S. Robinson, J. Vrba, P. Murphy, D. Rose, and C. Lowery, "Magnetoencephalographic recordings of visual evoked brain activity in the human fetus," *Lancet*, vol. 360, no. 9335, pp. 779–780, 2002.

[17] S. J. Paterson, J. H. Brown, M. K. Gsodl, M. H. Johnson, and A. Karmiloff-Smith, "Cognitive modularity and genetic disorders," *Science*, vol. 286, pp. 2355–2358, 1999.

[18] A. Karmiloff-Smith, "Development itself is the key to understanding developmental disorders," *Trends Cogn. Sci.*, pp. 389–398, 1998.

[19] J. Elman, E. A. Bates, M. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett, *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press, 1996.

[20] E. Bates, "The changing nervous system: Neurobehavioral consequences of early brain disorders," in *Plasticity, Localization and Language Development*. Oxford, U.K.: Oxford Univ. Press, 1997, pp. 214–253.

[21] M. I. Posner and S. E. Petersen, "The attention system of the human brain," *Annu. Rev. Neurosci.*, pp. 25–42, 1990.

[22] S. J. Paterson, S. Heim, J. T. Friedman, N. Choudhury, and A. A. Benasich, *Development of Structure and Function in the Infant Brain: Implications for Cognition, Language and Social Behaviour*. Oxford, U.K.: Elsevier, 2006, pp. 1087–1105.

[23] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini, "Developmental robotics: A survey," *Connect. Sci.*, vol. 15, no. 4, pp. 151–190, 2003.

[24] , J. A. S. Kelso, Ed., *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press/Bradford Books, 1995.

[25] P. Rochat, "Self-perception and action in infancy," *Exper. Brain Res.*, pp. 102–109, 1998.

[26] P. Rochat and T. Striano, "Perceived self in infancy," *Infant Behav. Develop.*, pp. 513–530, 2000.

[27] K. Doya, "Metalearning and neuromodulation," *Neural Netw.*, pp. 495–506, 2002.

[28] A. N. Meltzoff and M. K. Moore, "Imitation of facial and manual gestures by human neonates," *Science*, pp. 74–78, 1977.

[29] Y. Kuniyoshi and S. Sangawa, "Early motor development from. partially ordered neural-body dynamics: experiments with a. cortico-spinal-musculo-sleletal model," *Biol. Cybern.*, vol. 95, pp. 589–605, 2006.

[30] G. Taga, Y. Yamaguchi, and H. Shimizu, "Selforganized control of bipedal locomotion by neural oscillators in unpredictable environment," *Biol. Cybern.*, vol. 65, pp. 147–159, 1991.

[31] G. Taga, "Emergence of bipedal locomotion through entrainment among the neuromusculo-skeletal system and the environment," *Phys. D*, vol. 75, no. 1–3, pp. 190–208, 1994.

[32] H. Kimura, Y. Fukuoka, and K. Konaga, "Adaptive dynamic walking of a quadruped robot by using neural system model," *Adv. Robot.*, vol. 15, no. 8, pp. 859–876, 2001.

[33] R. Suzuki, I. Katsuno, and K. Matano, "Dynamics of neuron "ring"—Computer simulation of central nervous system of starfish," *Biol. Cybern.*, vol. 8, pp. 39–45, 1970.

[34] H. Sun and R. Jensen, "Body segment growth during infancy," *J. Biomech.*, vol. 21, no. 3, pp. 265–275, 1994.

[35] S. Ressler, Anthrokids—Anthropometric data of children 1977 [Online]. Available: http://www.itl.nist.gov/iaui/ovrt.projects/anthrokids/

[36] A. Freivalds, Incorporation of active elements into the articulated total body model Armstrong Aerospace Medical Research Lab., 1985, paper AAMRL-TR-85-061.

[37] S. Goodall, J. Reggia, Y. Chen, E. Ruppin, and C. Whitney, "A computational model of acute focal cortical lesions," *Stroke*, vol. 28, pp. 101–109, 1997.

[38] E. M. Izhikevich and G. M. Edelman, "Large-scale model of mammalian thalamocortical systems," *Proc. Nat. Acad. Sci.*, vol. 105, no. 9, pp. 3593–3598, 2008.

[39] K. Kinjo, C. Nabeshima, S. Sangawa, and Y. Kuniyoshi, "A neural model for exploration and learning of embodied movement patterns," *J. Robot. Mechatron.*, vol. 20, no. 3, pp. 358–366, 2008.

[40] A. Pitti, H. Alirezaei, and Y. Kuniyoshi, "Cross-modal and scale-free action representations through enaction," *Neural Netw.*, 2009.

[41] K. Hosoda, H. Takayama, and T. Takuma, "Bouncing monopod with bio-mimetic muscular-skeleton system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. 2008 (IROS'08)*, 2008.

[42] L. Righetti and A. J. Ijspeert, "Design methodologies for central pattern generators: An application to crawling humanoids," *Proc. Robot., Sci. Syst.*, pp. 191–198, 2006.

[43] S. Degallier, L. Righetti, L. Natale, N. Nori, G. Metta, and A. Ijspeert, "A modular, bio-inspired architecture for movement generation for the infant-like robot icub," in *Proc. 2nd IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatron. (BioRob)*, 2008.

[44] S. Schaal, D. Sternad, R. Osu, and M. Kawato, "Rhythmic arm movement is not discrete," *Nat. Neurosci.*, vol. 7, no. 10, pp. 1137–1144, 2004.

[45] G. Buzsaki, *Rhythms of the Brain*. Oxford, U.K.: Oxford Univ. Press, 2006.

[46] T. Flash and T. J. Sejnowski, "Computational approaches to motor control," *Curr. Opinion Neurobiol.*, vol. 11, pp. 655–662, 2001.

[47] D. M. Amodio and C. D. Frith, "Meeting of minds: The medial frontal cortex and social cognition," *Nat. Rev. Neurosci.*, vol. 7, pp. 268–277, 2006.

[48] S.-J. Blakemore and U. Frith, *The Learning Brain: Lessons for Education*. Oxford, U.K.: Blackwell, 2005.

[49] H. Wagner and R. Blickhan, "Stabilizing function of skeletal muscles: An analytical investigation," *J. Theor. Bilo.*, pp. 163–179, 1999.

[50] R. M. Alexander, "Tendon elasticity and muscle function," *Comp. Biochem. Physiol. A, Mol. Integr. Physiol.*, vol. 133, no. 4, pp. 1001–1011, Dec. 2002.

[51] R. Blickhan, "The spring-mass model for running and hopping," *J. Biomechan.*, vol. 12, no. 11–12, pp. 1217–1227, 1989.

[52] D. E. Koditscheck and M. Buehler, "Analysis of a simplified hopping robot," *Int. J. Robot. Res.*, vol. 10, pp. 269–281, 1991.

[53] M. Ahmadi and M. Buehler, "Stable control of a simulated one-legged running robot with hip and leg compliance," *IEEE Trans. Robot. Autom.*, vol. 13, pp. 96–104, 1997.

[54] A. Seyfarth, H. Geyer, M. Gunther, and R. Blickhan, "A movement criterion for running," *J. Biomech.*, vol. 35, pp. 649–655, 2002.

[55] M. Raibert, *Legged Robots That Balance*. Cambridge, MA: MIT Press, 1986.

[56] H. Geyer, A. Seyfarth, and R. Blickhan, "Compliant leg behaviour explains basic dynamics of walking and running," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 273, pp. 2861–2867, 2006.

[57] F. Iida, J. Rummel, and A. Seyfarth, "Bipedal walking and running with compliant legs," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2007, pp. 3970–3975.

[58] S. H. Hyon and T. Mita, "Development of a biologically inspired hopping robot—Kenken," in *Int. Conf. Robot. Autom.*, May 2002, pp. 3984–3991.

[59] J. W. Hurst, J. E. Chestnutt, and A. A. Rizzi, "Design and philosophy of the bimasc, a highly dynamic biped," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2007, pp. 1863–1868.

[60] B. Vanderborght, R. Van Ham, B. Verrelst, M. Van Damme, and D. Lefeber, "Overview of the lucy-project: Dynamic stabilisation of a biped powered by pneumatic artificial muscles," *Adv. Robot.*, vol. 22, no. 10, pp. 1027–1051, 2008.

[61] K. Hosoda and T. Ueda, "Contribution of external rotation to emergence of biped walking," in *Proc. Int. Symp. Adapt. Motion Animals Machines*, 2008.

[62] K. Narioka and K. Hosoda, "Designing synergictic walking of a whole-body humanoid driven by pneumatic artificial muscles," *Adv. Robot.*, vol. 22, no. 10, pp. 1107–1123, 2008.

[63] K. Hosoda and K. Narioka, "Synergistic 3d limit cycle walking of an anthropomorphic biped robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 470–475.

[64] R. Niiyama and Y. Kuniyoshi, "A pneumatic biped with an artificial musculoskeletal system," in *Proc. 4th Int. Symp. Adapt. Motion Animals Machines (AMAM 2008)*, 2008.

[65] T. Takuma, S. Hayashi, and K. Hosoda, "3d bipedal robot with tunable leg compliance mechanism for multi-modal locomotion," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. 2008 (IROS'08)*.

[66] R. Pfeifer, F. Iida, and G. Gömez, "Morphological computation for adaptive behavior and cognition," *Int. Congr. Series*, vol. 1291, pp. 22–29, 2006.

[67] T. McGeer, "Passive walking with knees," in *Proc. 1990 IEEE Int. Conf. Robot. Autom.*, 1990.

[68] T. Minato, Y. Yoshikawa, T. Noda, S. Ikemoto, H. Ishiguro, and M. Asada, "Cb^2: A child robot with biomimetic body for cognitive developmental robotics," in *Proc. IEEE/RSJ Int. Conf. Human. Robots*, 2007.

[69] H. Ishiguro, T. Ono, M. Imai, T. Kanda, and R. Nakatsu, "Robovie: An interactive humanoid robot," *Int. J. Ind. Robot*, vol. 28, no. 6, pp. 498–503, 2001.

[70] H. Kozima, "Infanoid: A babybot that explores the social environment," in *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, K. Dautenhahn, A. H. Bond, L. Canamero, and B. Edmonds, Eds. Amsterdam, The Netherlands: Kluwer Academic, 2002, pp. 157–164.

[71] M. Hackel, S. Schwope, J. Fritsch, B. Wrede, and G. Sagerer, "A humanoid robot platform suitable for studying embodied interaction," in *Proc. 2005 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2005, pp. 56–61.

[72] D. Vernon, G. Metta, and G. Sandini, "The icub cognitive architecture: Interactive development in a humanoid robot," in *Proc. 6th IEEE Int. Conf. Develop. Learn.*, 2007.

[73] T. Miyashita, T. Tajika, H. Ishiguro, K. Kogure, and N. Hagita, "Haptic communication between humans and robots," in *Proc. 12th Int. Symp. of Robot. Res.*, 2005.

[74] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent ASIMO: System overview and integration," in *Proc. 2002 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2002, pp. 2478–2483.

[75] M. Fujita, Y. Kuroki, T. Ishida, and T. Doi, "Autonomous behavior control architecture of entertainment humanoid robot sdr-4x," in *Proc. 2003 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2003, pp. 960–967.

[76] T. Hashimoto, S. Hiramatsu, T. Tsuji, and H. Kobayashi, "Development of the face robot saya for rich facial expressions," in *Proc. SICE-ICASE Int. Joint Conf.*, 2006, pp. 5423–5428.

[77] G. Cheng, S.-H. Hyon, J. Morimoto, A. Ude, J. G. Hale, G. Colvin, W. Scroggin, and S. C. Jacobsen, "Cb: A humanoid research platform for exploring neuroscience," *Adv. Robot.*, vol. 21, no. 10, pp. 1097–1114, 2007.

[78] T. Minato, M. Shimada, H. Ishiguro, and S. Itakura, "Development of an android robot for studying human-robot interaction," in *Proc. 17th Int. Conf. Ind. Eng. Applicat. Artif. Intell. Expert Syst.*, Ottawa, ON, Canada, 2004, pp. 424–434.

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

[80] D. Sakamoto, T. Kanda, T. Ono, H. Ishiguro, and N. Hagita, "Android as a telecommunication medium with human like presence," in *Proc. 2nd ACM/IEEE Int. Conf. Human-Robot Interact.*, 2007.

[81] S. Ikemoto, T. Minato, and H. Ishiguro, "Analysis of physical human-robot interaction for motor learning with physical help," in *Proc. IEEE/RSJ Int. Conf. Human. Robots*, 2008.

[82] M. I. Stamenov, "Body image and body schema," in *Body Schema, Body Image, and Mirror Heurons*. Amsterdam, The Netherlands: John Benjamins, 2005, pp. 22–43.

[83] V. S. Ramachandran and S. Blakeslee*, Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York: Harper Perennial, 1998.

[84] K. Hosoda and M. Asada, "Versatile visual servoing without knowledge of true Jacobian," *Proc. IROS'94*, pp. 186–193, 1994.

[85] S. G. D. Bullock and F. H. Guenther, "A self-organized neural model of motor equivalent reaching and tool use by a multijoint arm," *J. Cogn. Neurosci.*, vol. 5, no. 4, pp. 408–435, 1993.

[86] C. G. Sun and B. Scassellati, "A fast and efficient model for learning to reach," *Int. J. Human. Robot.*, vol. 2, no. 4, pp. 391–414, 2005.

[87] A. Iriki, M. Tanaka, S. Obayashi, and Y. Iwamura, "Self-images in the video monitor coded by monkey intraparietal neurons," *Neurosci. Res.*, vol. 40, pp. 163–173, 2001.

[88] F. Sawa, M. Ogino, and M. Asada, "Body image constructed from motor and tactle images with visual informaiton," *Int. J. Human. Robot.*, vol. 4, pp. 347–364, 2007.

[89] A. N. Meltzoff and M. K. Moore, "Explaining facial imitation: A theoretical model," *Early Develop. Parent.*, pp. 179–192, 1997.

[90] C. Breazeal, D. Buchsbaum, J. Gray, D. Gatenby, and B. Blumberg, "Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots," *Artif. Life*, vol. 11, pp. 1–32, 2005.

[91] C. Nabeshima, M. Lungarella, and Y. Kuniyoshi, "Timing-based model of body schema adaptation and its role in perception and tool use: A robot case study," in *Proc. 4th Int. Conf. Develop. Learn. (ICDL'05)*, Osaka, Japan, Jul. 2005, pp. 7–12.

[92] Y. Yoshikawa, H. Kawanishi, M. Asada, and K. Hosoda, "Body scheme acquisition by cross map learning among tactile, image, and proprioceptive spaces," in *Proc. 2nd Int. Workshop Epigen. Robot.: Model. Cogn. Develop. Robot. Syst.*, 2002, pp. 181–184.

[93] Y. Yoshikawa, "Subjective robot imitation by finding invariance," Ph.D. dissertation, Osaka Univ., Osaka, Japan, 2005.

[94] A. Stoytchev, "Toward video-guided robot behaviors," in *Proc. 7th Int. Conf. Epigen. Robot.*, 2007, pp. 165–172.

[95] M. Hersch, E. Sauser, and A. Billard, "Online learning of the body schema," *Int. J. Human. Robot.*, vol. 5, no. 2, pp. 161–181, 2008.

[96] M. Hikita, S. Fuke, M. Ogino, T. Minato, and M. Asada, "Visual attention by saliency leads cross-modal body representation," in *Proc. 7th Int. Conf. Develop. Learn. (ICDL'08)*, 2008.

[97] L. Itti and F. Pighin, "Realistic avatar eye and head animation using a neurobiological model of visual attention," in *Proc. SPIE 48th Annu. Int. Symp. Opt. Sci. Technol.*, 2003, vol. 5200, pp. 64–78.

[98] A. Iriki, M. Tanaka, and Y. Iwamura, "Coding of modified body schema during tool use by macaque postcentral neurones," *Cogn. Neurosci. Neuropsychol.*, vol. 7, no. 14, pp. 2325–2330, 1996.

[99] S. Fuke, M. Ogino, and M. Asada, "Vip neuron model:head-centered cross-modal representation of the peri-personal space around the face," in *Proc. 7th IEEE Int. Conf. Develop. Learn.*, 2008, pp. 145–150.

[100] J. R. Duhamel, C. L. Colby, and M. E. Goldberg, "Ventral intraparietal area of the macaque: Congruent visual and somatic response properties," *J. Neurophysiol.*, vol. 79, pp. 126–136, 1998.

[101] M. S. A. Graziano and D. F. Cooke, "Parieto-frontal interactions, personal space, and defensive behavior," *Neuropsychologia*, vol. 44, pp. 845–859, 2006.

[102] M. I. Sereno and R. Huang, "A human parietal face area contains aligned head-centered visual and tactile maps," *Nature Neurosci.*, vol. 9, pp. 1337–1343, 2006.

[103] R. A. Andersen, "Encoding of intention and spatial location in the posterior parietal cortex," *Cerebral Cortex*, vol. 5, pp. 457–469, 1995.

[104] G. J. Bratt, R. A. Andersen, and J. R. Stoner, "Visual receptive field organization and cortico-cortical connections of the lateral intraparietal are (area lip) in the macaque," *J. Comp. Neurol.*, vol. 299, pp. 421–445, 1990.

[105] O. A. Mullette-Gillman, Y. E. Cohen, and J. M. Groh, "Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus," *J. Neurophysiol.*, vol. 94, no. 4, pp. 2331–2352, 2005.

[106] K. Ogawa and T. Inui, "Lateralization of the posterior parietal cortex for internal monitoring of self-versus externally generated movements," *J. Cogn. Neurosci.*, vol. 19, pp. 1827–1835, 2007.

[107] K. Ogawa, T. Inui, and T. Sugio, "Separating brain regions involved in internally guided and visual feedback control of moving effectors: An event-related fMRI study," *Neuroimage*, vol. 32, no. 4, pp. 1760–1770, 2006.

[108] T. Inui, “A theory of image generation: Normal and pathological cases,” *Gendaishisou*, vol. 35, no. 6, pp. 233–245, 2007, (in Japanese).

[109] C. Breazeal and B. Scassellati, “Infant-like social interactions between a robot and a human caregiver,” *Adapt. Behav.*, vol. 8, no. 1, pp. 49–74, 2000.

[110] P. Rochat, J. G. Querido, and T. Striano, “Emerging sensitivity to the timing and structure of protoconversation in early infancy,” *Develop. Psychol.*, vol. 35, no. 4, pp. 950–957, 1999.

[111] P. Rochat, *The Infant’s World*. Cambridge, MA: Harvard Univ. Press, 2001, ch. 4.

[112] M. Ogino, T. Ooide, A. Watanabe, and M. Asada, “Acquiring peekaboo communication: Early communication model based on reward prediction,” in *Proc. 6th IEEE Int. Conf. Develop. Learn.*, 2007, pp. 116–121.

[113] R. Paz, J. G. Pelletier, E. P. Bauer, and D. Pare, “Emotion enhancement of memory via amygdala-driven facilitation of rhinal interactions,” *Nature Neurosci.*, vol. 9, no. 10, pp. 1321–1329, 2006.

[114] J. L. McGaugh, *Memory and Emotion*. London, U.K.: Orion, 2003.

[115] W. Schultz, P. Dayan, and P. F. Strick, “A neural substrate of prediction and reward,” *Science*, vol. 275, pp. 236–250, 1997.

[116] N. A. Mirza, C. L. Nehaniv, K. Dautenhahn, and R. te Boekhorst, “Grounded sensorimotor interaction histories in an information theoretic metric space for robot ontogeny,” *J. Adapt. Behav.*, vol. 15, pp. 167–187, 2007.

[117] J. Piaget, *The Construction of Reality in the Child*. New York: Basic, 1954.

[118] B. Elsner, “Infants’ imitation of goal-directed actions: The role of movements and action effects,” *Acta Psychol.*, vol. 124, pp. 44–59, 2007.

[119] J. A. Sommerville, A. L. Woodward, and A. Needham, “Action experience alters 3-month-old infants’ perception of others’ actions,” *Cognition*, vol. B1–11, 1996.

[120] G. Rizzolatti, L. Fogassi, and V. Gallese, “Neurophysiological mechanisms underlying the understanding and imitation of action,” *Nature Rev. Neurosci.*, vol. 2, pp. 661–670, 2001.

[121] K. Ogawa and T. Inui, “Neural basis of hierarchical action representations for imitation: An fMRI study,” in *38th Annu. Meeting Soc. Neurosci.*, 2008.

[122] C. Grefkes and G. R. Fink, “The functional organization of the intraparietal sulcus in humans and monkeys,” *J. Anatomy*, vol. 207, pp. 3–17, 2005.

[123] D. M. Wolpert, S. J. Goodbody, and M. Husain, “Maintaining internal representations: The role of the human superior parietal lobe,” *Nature Neurosci.*, vol. 1, pp. 529–533, 1998.

[124] L. Moll and H. G. Kuypers, “Premotor cortical ablations in monkeys: Contralateral changes in visually guided reaching behavior,” *Science*, vol. 198, pp. 317–319, 1977.

[125] S. S. Jones, “Imitation in infancy—the development of mimicry,” *Psychol. Sci.*, vol. 18, no. 7, pp. 593–599, 2007.

[126] J. F. Werker and R. C. Tees, “Cross-language speech perception: Evidence for perceptual reorganization during the first year of life,” *Infant Behav. Develop.*, vol. 25, pp. 121–133, 2002.

[127] P. K. Kuhl and A. N. Meltzoff, “Infant vocalizations in response to speech: Vocal imitation and developmental change,” *J. Acoust. Soc. Amer.*, vol. 100, pp. 2415–2438, 1996.

[128] C. T. Sasaki, P. A. Levine, I. T. Laitman, and E. S. Crelin, “Postnatal developmental descent of the epiglottis in man,” *Arch. Otolaryngol.*, vol. 103, pp. 169–171, 1977.

[129] G. Dehaene-Lambertz, S. Dehaene, and L. Hertz-Pannier, “Functional neuroimaging of speech perception in infants,” *Science*, vol. 298, pp. 2013–2015, 2002.

[130] J. Gervain, F. Macagno, S. Cogoi, M. Penä, and J. Mehler, “The neonate brain detects speech structure,” *Proc. Nat. Acad. Sci. USA*, vol. 105, pp. 14222–14227, 2008.

[131] F. H. Guenther, S. S. Ghosh, and J. A. Tourville, “Neural modeling and imaging of the cortical interactions underlying syllable production,” *Brain Lang.*, vol. 96, pp. 280–301, 2006.

[132] G. Westermann and E. R. Miranda, “A new model of sensorimotor coupling in the development of speech,” *Brain Lang.*, vol. 89, pp. 393–400, 2004.

[133] H. Kanda, T. Ogata, K. Komatani, and H. G. Okuno, “Segmenting acoustic signal with articulatory movement using recurrent neural network for phoneme acquisition,” in *Proc. 2008 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2008, pp. 1712–1717.

[134] J. Hörnstein and J. Santos-Victor, “A unified approach to speech production and recognition based on articulatory motor representations,” in *Proc. 2007 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2007, pp. 3442–3447.

[135] T. Kokkinaki and G. Kugiumutzakis, “Basic aspects of vocal imitation in infant-parent interaction during the first 6 months,” *J. Reproduct. Infant Psychol.*, vol. 18, pp. 173–187, 2000.

[136] N. Masataka and K. Bloom, “Accoustic properties that determine adult’s preference for 3-month-old infant vocalization,” *Infant Behav. Develop.*, vol. 17, pp. 461–464, 1994.

[137] M. Pélaez-Nogueras, J. L. Gewirtz, and M. M. Markham, “Infant vocalizations are conditioned both by maternal imitation and motherese speech,” *Infant Behav. Develop.*, vol. 19, p. 670, 1996.

[138] B. de Boer, “Self organization in vowel systems,” *J. Phonetics*, vol. 28, no. 4, pp. 441–465, 2000.

[139] P.-Y. Oudeyer, “The self-organization of speech sounds,” *J. Theor. Biol.*, vol. 233, no. 3, pp. 435–449, 2005.

[140] Y. Yoshikawa, J. Koga, M. Asada, and K. Hosoda, “A constructivist approach to infants’ vowel acquisition through mother-infant interaction,” *Connect. Sci.*, vol. 15, no. 4, pp. 245–258, 2003.

[141] K. Miura, Y. Yoshikawa, and M. Asada, “Unconscious anchoring in maternal imitation that helps finding the correspondence of caregiver’s vowel categories,” *Adv. Robot.*, vol. 21, pp. 1583–1600, 2007.

[142] H. Ishihara, Y. Yoshikawa, K. Miura, and M. Asada, “Caregiver’s sensorimotor magnets lead infant’s vowel acquisition through auto mirroring,” in *Proc. 7th IEEE Int. Conf. Develop. Learn.*, 2008.

[143] P. K. Kuhl, “Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not,” *Percept. Psychophys.*, vol. 50, pp. 93–107, 1991.

[144] P.-Y. Oudeyer, “Phonemic coding might result from sensory-motor coupling dynamics,” in *Proc. 7th Int. Conf. Simul. Adapt. Behav. (SAB02)*, 2002, pp. 406–416.

[145] K. Miura, Y. Yoshikawa, and M. Asada, “Realizing being imitated: Vowel mapping with clearer articulation,” in *Proc. 7th IEEE Int. Conf. Develop. Learn.*, 2008.

[146] S. Baron-Cohen, *Mindblindness*. Cambridge, MA: MIT Press, 1995.

[147] *Joint Attention: It’s Origins and Role in Development*, C. Moore and P. Dunham, Eds. New York: Lawrence Erlbaum, 1995.

[148] M. H. Johnson, S. Dziurawiec, H. Dllis, and J. Morton, “Newborns’ preferential tracking of face-like stimuli and its subsequent decline,” *Cognition*, vol. 40, pp. 1–19, 1991.

[149] M. H. Johnson, “Subcortical face processing,” *Nature Rev.*, vol. 6, pp. 766–774, Oct. 2005.

[150] , A. Slater and M. Lewis, Eds., *Introduction to Infant Development*. Oxford, U.K.: Oxford Univ. Press, 2007.

[151] I. Fasel, N. Butko, and J. Movellan, “Modeling the embodiment of early social development and social interaction: Learning about human faces during the first six minutes of life,” in *Proc. Soc. Res. Child Develop Bien. Meeting*, 2007.

[152] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Choen, “Defining and quantifying the social phenotype in autism,” *Social Phenotype Autism*, vol. 159, pp. 895–908, 2002.

[153] O. I. Lovaas, “Behavioral treatment and normal educational and intellectual functioning in young antistic children,” *J. Consult. Clin. Psychol.*, vol. 55, no. 1, pp. 3–9, 1987.

[154] F. Shic, B. Scassellati, D. Lin, and K. Chawarska, “Measuring context: The gaze patterns of children with autism evaluated from the bottom-up,” in *Proc. 6th IEEE Int. Conf. Develop. Learn.*, 2007.

[155] M. Ogino, A. Watanabe, and M. Asada, “Detection and categorization of facial image through the interaction with caregiver,” in *Proc. 7th Int. Conf. Develop. Learn. (ICDL’08)*, 2008.

[156] N. Otsu and T. Kurita, “A new scheme for practical flexible and intelligent vision systems,” in *Proc. IAPR Workshop Comput. Vision*, 1988, pp. 431–435.

[157] M. Tomasello, “Joint attention: It’s origins and role in development,” in *Joint Attention as Social Cognition*, C. Moore and P. Dunham, Eds. New York: Lawrence Erlbaum, 1995, pp. 103–130.

[158] T. Farroni, G. Csibra, F. Simion, and M. H. Johnson, “Eye contact detection in humans from birth,” *Proc. Nat. Acad. Sci. USA*, vol. 99, pp. 9602–9605, 2002.

[159] P. Mundy, J. Card, and N. Fox, “EEG correlates of the development of infant joint attention skill,” *Develop. Psychol.*, vol. 36, pp. 325–338, 2000.

[160] B. Scassellati, “Computational for metaphors, analogy, and agents,” in *Imitation and Mechanism of Joint Attention: A Developmental Structure for Building Social Skills on a Human. Robot*, C. L. Nehaniv, Ed. Berlin, Germany: Springer-Verlag, 1999, pp. 176–195.

[161] H. Kozima, C. Nakagawa, and H. Yano, “Attention coupling as a prerequisite for social interaction,” in *Proc. IEEE Int. Workshop Robot Human Interact. Commun.*, 2003, pp. 109–114.

[162] Y. Nagai, M. Asanda, and K. Hosoda, "Learning for joint attention helped by functional development," *Adv. Robot.*, vol. 20, no. 10, p. 1165, 2006.

[163] Y. Nagai, K. Hosoda, A. Morita, and M. Asada, "A constructive model for the development of joint attention," *Connect. Sci.*, vol. 15, pp. 211–229, 2003.

[164] J. Triesch, G. Teuscher, G. Deak, and E. Carlson, "Gaze following: Why (not) learn it," *Develop. Sci.*, vol. 9, no. 2, pp. 125–147, 2006.

[165] F. Kaplan and V. Hafner, "The challenges of joint attention," *Interact. Studies*, vol. 7, no. 2, pp. 135–169, 2006.

[166] D. Baldwin, "Infants' contribution to the achievement of joint reference," *Child Develop.*, vol. 62, pp. 875–890, 1991.

[167] H. Sumioka, Y. Yoshikawa, and M. Asada, "Development of joint attention related actions based on reproducing interaction contingency," in *Proc. 7th IEEE Int. Conf. Develop. Learn.*, 2008.

[168] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, 2000.

[169] S. Blakemore, P. Boyer, M. Pachot-Clouard, A. Meltzoff, C. Segebarth, and J. Decety, "The detection of contingency and animacy from simple animations in the human brain," *Cerebral Cortex*, vol. 13, no. 8, pp. 837–844, 2003.

[170] E. Bates, P. Dale, and D. Thal, "Individual differences and their implications for theories of language development," in *Handbook of Child Language*, Fletcher and MacWhinney, Eds. Oxford, U.K.: Basil Blackwell, 1995, pp. 96–151.

[171] D. Roy and A. Pentland, "Learning words from sights and sounds: A computational model," *Cogn. Sci.*, vol. 26, pp. 113–146, 2002.

[172] C. Yu, L. Smith, Krystal, A. Klein, and R. Shiffrin, "Hypothesis testing and associative learning in cross-situational word learning: Are they one and the same?," in *Proc. 29th Annu. Conf. Cogn. Sci. Soc.*, 2007, pp. 737–742.

[173] C. Yu, D. Ballard, and R. Aslin, "The role of embodied intention in early lexical acquisition," *Cogn. Sci.*, 2005.

[174] Y. Yoshikawa, T. Nakano, M. Asada, and H. Ishiguro, "Multimodal joint attention through cross facilitative learning based on μx principle," in *Proc. 7th IEEE Int. Conf. Develop. Learn.*, 2008.

[175] A. Mehrabian, *Implicit Communication of Emotions and Attitudes*. London, U.K.: Wadsworth, 1981.

[176] M. H. Johnson and J. Morton, *Biology and Cognitive Development: The Case of Face Recognition*. London, U.K.: Blackwell, 1991.

[177] D. Rosenstein and H. Oster, *Differential Facial Responses to Four Basic Tastes in Newborns*. London, U.K.: Blackwell, 1988, vol. 59, pp. 1555–1568.

[178] D. Matsui, T. Minato, K. F. MacDorman, and H. Ishiguro, "Generating natural motion in an android by mapping human motion," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2005, pp. 1089–1096.

[179] T. Hashimoto, M. Sennda, and H. Kobayashi, "Realization of realistic and rich facial expressions by face robot," in *Proc. 2004 1st IEEE Techn. Exhib. Based Conf. Robot. Autom.*, Nov. 2004, pp. 37–38.

[180] H. Papousek and M. Papousek, "Intuitive parenting: A dialectic counterpart to the infant's precocity in integrative capacities," in *Handbook of Infant Development*. London, U.K.: Blackwell, 1987, pp. 669–720.

[181] G. Gergely and J. S. Watson, "Early socio-emotional development: Contingency perception adn the social-biofeedback model," in *Early Social Cognition: Understanding Others in the First Months of Life*, P. Rochat, Ed. Mahwah, NJ: Lawrence Erlbaum, 1999, pp. 101–136.

[182] A. Watanabe, M. Ogino, and M. Asada, "Mapping facial expression to internal states based on intuitive parenting," *J. Robot. Mechatron.*, vol. 19, no. 3, pp. 315–323, 2007.

[183] T. Singer, B. Seymour, J. O'Doherty, H. Kaube, R. J. Dolan, and C. D. Frith, "Empathy for pain involves the affective but not sensory components of pain," *Science*, vol. 303, no. 20, pp. 1157–1162, 2004.

[184] K. D. Pruett, *Me, Myself and I: How Children Build Their Sense of Self-18 to 36 Months*. New York: Goddard, 1999.

[185] H. Asoh, S. Akaho, O. Hasegawa, T. Yoshimura, and S. Hayamizu, "Intermodal learning of multimodal interaction systems," in *Proc. Int. Workshop Human Interface Technol.*, 1997.

[186] K. Ishiguro, N. Otsu, and Y. Kuniyoshi, "Inter-modal learnig and object concept acquisition," in *Proc. IAPR Conf. Machine Vision Applicat. (MVA2005)*, 2005.

[187] L. Steels and F. Kaplan, "Aibo's first words. The social learning of language and meaning," *Evol. Commun.*, vol. 4, no. 1, pp. 3–31, 2001.

[188] N. Iwahashi, "Language acquisition through a human-robot interface by combining speech, visual, and behavioral information," *Inf. Sci.*, vol. 156, pp. 109–121, 2003.

[189] E. M. Markman, *Categorization in Children: Problems of Induction*. Cambridge, MA: MIT Press/Bradford Books, 1989.

[190] B. Landau, L. B. Smith, and S. Jones, "The importance of shape in early lexical learning," *Cogn. Develop.*, vol. 3.

[191] D. K. Roy, "Learning visually-grounded words and syntax for a scene description task," *Comput. Speech Lang.*, vol. 16, pp. 353–385, 2002.

[192] A. Toyomura and T. Omori, "A computational model for taxonomy-based word learning inspired by infant developmental word acquisition," *IEICE Inf. Syst.*, vol. 88, no. 10, pp. 2389–2398, 2005.

[193] D. Walther, U. Rutishauser, C. Koch, and P. Perona, "Selective visual attention enables learning and recognition of multiple objects in cluttered scenes," *Comput. Vision Image Understand.*, vol. 100, pp. 41–63, 2005.

[194] M. Ogino, M. Kikuchi, and M. Asada, "Active lexicon acquisition based on curiosity," in *Proc. 5th Int. Conf. Develop. Learn.*, 2006.

[195] P. F. Dominey, M. Hoen, and T. Inui, "A neurolinguistic model of grammatical construction processing," *J. Cogn. Neurosci.*, pp. 1–20, 2006.

[196] P. F. Dominey, T. Inui, and M. Hoen, "Neural network processing of natural language: Towards a unified model of corticostriatal function in learning sentence comprehension and non-linguistic sequencing," *Brain Lang.*, 2008.

[197] M. Dapretto and S. Bookheimer, "Form and content: Dissociating syntax and semantics in sentence comprehension," *Neuron*, vol. 24, no. 2, pp. 427–432, 1999.

[198] A. D. Friederici, S. A. Rueschemeyer, A. Hahne, and C. J. Fiebach, "The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes," *Cerebral Cortex*, vol. 13, no. 2, pp. 170–177, 2003.

[199] T. Inui, K. Ogawa, and M. Ohba, "Role of left inferior frontal gyrus in the processing of particles," *NeuroReport*, vol. 18, no. 5, pp. 431–434, 2007, (in Japanese).

[200] K. Ogawa, M. Ohba, and T. Inui, "Neural basis of syntactic processing of simple sentences ," *NeuroReport*, vol. 18, no. 14, pp. 1437–1441, 2007, (in Japanese).

[201] D. Caplan, *Language: Structure, Processing, and Disorders*. Cambridge, MA: MIT Press, 1992.

[202] J. R. Saffran, R. N. Aslin, and E. L. Newport, "Statistical learning by 8-month-old infants," *Science*, vol. 274, no. 5294, pp. 1926–1928, 1996.

[203] A. S. Reber, "Implicit learning of artificial grammars," *J. Verbal Learn. Verbal Behav.*, vol. 6, pp. 855–863, 1967.

[204] K. Hirsh-Pasek and R. M. Golinkoff, *The Origins of Grammar: Evidence from Early Language Comprehension*. Boston, MA: MIT Press, 1996.

[205] K. Narioka, R. Niiyama, K. Hosoda, and Y. Kuniyoshi, "A baby robot with an artificial musculoskeletal system," in *Proc. 26th Annu. RSJ Meetings*, 2008, vol. 1J2-01, (in Japanese).

[206] S. Hosaka, C. Yoshida, Y. Kuniyoshi, and M. Asada, "Measurement of mother-infant interaction using tactile sensor suits," in *Proc. 8th Annu. Baby Sci. Meetings*, 2008, (in Japanese).

[207] C. Higgins, J. Campos, and R. Kermoian, "Effects of self-produced locomotion on infant postural compensation to optic flow," *Develop. Psychol.*, vol. 32, pp. 836–841, 1996.

[208] L. Natale, F. Orabona, G. Metta, and G. Sandini, "Exploring the world through grasping: A developmental approach," in *Proc. 6th CIRA Symp.*, 2005.

[209] L. Natale, F. Orabona, G. Metta, and G. Sandini, "Sensorimotor coordination in a 'baby; robot: Learning about objects through grasping," *Progr. Brain Res.: From Action to Cogn.*, vol. 164, 2007.

[210] L. Natale, F. Nori, and G. Metta, "Learning precise 3d reaching in a humanoid robot," in *Proc. 6th IEEE Int. Conf. Develop. Learn.*, 2007.

[211] Y. Kuniyoshi, M. Inaba, and H. Inoue, "Learning by watching," *IEEE Trans. Robot. Autom.*, vol. 10, pp. 799–822, 1994.

[212] S. Schaal, "Is imitation learning the route to humanoid robots?," *Trends Cogn. Sci.*, pp. 233–242, 1999.

[213] , K. Dautenhahn and C. L. Nehaniv, Eds., *Imitation in Animals and Artifacts*. Cambridge, MA: MIT Press, 2002.

Minoru Asada (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.

In April 1995, he became a Professor of the Osaka University. Since April 1997, he has been a Professor of the Department of Adaptive Machine Systems at the Graduate School of Engineering, Osaka University. From August 1986 to October 1987, he was a Visiting Researcher of Center for Automation Research, University of Maryland, College Park, MD.

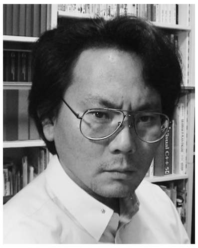
Dr. Asada received many awards including the Best Paper Award of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS92) and the Commendation by the Minister of Education, Culture, Sports, Science and Technology, Japanese Government as Persons of Distinguished Services to enlightening people on science and technology. He was the President of the International RoboCup Federation (2002–2008). Since 2005, he has been the Research Director of "ASADA Synergistic Intelligence Project" of ERATO (Exploratory Research for Advanced Technology by Japan Science and Technology Agency).

Koh Hosoda (M'93) received the Ph.D. degree in mechanical engineering from Kyoto University, Japan, in 1993.

From 1993 to 1997, he was a Research Associate of Mechanical Engineering for Computer-Controlled Machinery, Osaka University. Since February 1997, he has been an Associate Professor of the Department of Adaptive Machine Systems, Osaka University. Since November 2005, he has been a Group Leader of the JST Asada ERATO Project, as well.

Yasuo Kuniyoshi (M'02) received M.Eng. and Ph.D. degrees in information technology from the University of Tokyo, Japan, in 1988 and 1991, respectively.

He is a Professor at the Department of Mechano-Informatics, School of Information Science and Technology, The University of Tokyo, Japan. From 1991 to 2000, he was a Research Scientist and then a Senior Research Scientist at Electrotechnical Laboratory, AIST, MITI, Japan. From 1996 to 1997, he was a Visiting Scholar at MIT AI Lab. In 2001, he was appointed as an Associate Professor at the University of Tokyo. Since 2005, he has been a Professor at the same university. Since November 2005, he has been a Group Leader of the JST Asada ERATO Project. His research interests include emergence and development of embodied cognition, humanoid robot intelligence, machine understanding of human actions and intentions. He published over 400 technical papers and received IJCAI 93 Outstanding Paper Award, Best Paper Awards from Robotics Society of Japan, Sato Memorial Award for Intelligent Robotics Research, Okawa Publications Prize, Tokyo Techno Forum 21 Gold Medal Award and other awards. For further information about his research, visit http://www.isi.imi.i.u-tokyo.ac.jp

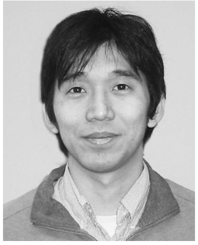


Hiroshi Ishiguro (M'90) received the D.Eng. degree in systems engineering from the Osaka University, Japan in 1991.

He is currently Professor in the Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University. Since 2002, he has also been a Visiting Group Leader of the Intelligent Robotics and Communication Laboratories at the Advanced Telecommunications Research Institute, where he previously worked as Visiting Researcher (1999–2002). He was previously Research Associate (1992–1994) in the Graduate School of Engineering Science at Osaka University and Associate Professor (1998–2000) in the Department of Social Informatics at Kyoto University. He was also Visiting Scholar (1998–1999) at the University of California, San Diego. He then became Associate Professor (2000–2001), Professor (2001–2002) in the Department of Computer and Communication Sciences at Wakayama University, and Professor (2002–2009) in the Department of Adaptive Machine Systems in the Graduate School of Engineering at Osaka University. His research interests include distributed sensor systems, interactive robotics, and android science.

Toshio Inui received the Ph.D. degree in psychology from Kyoto University, Kyoto, Japan, in 1985.

He is now a Professor at the Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University. He is also the Leader of Synergistic Intelligence Mechanism Group in ERATO Asada Synergistic Intelligence Project. His majors are cognitive science, cognitive neuroscience and computational neuroscience. Currently, he is engaged in research of neural basis of cognitive development, verbal, and nonverbal communication. He is an executive committee member of the Neuropsychology Association of Japan, the Japanese Society for Cognitive Psychology, the Japanese Neuro-ophthalmology Society, and the Japan Human Brain Mapping Society. He serves on the editorial board of *Neural Networks*. His publications include Inui, T., and McClelland, J. L. (Eds., 1996) *Attention and Performance XVI : Information Integration in Perception and Communication* (Cambridge, MA: The MIT Press).



Yuichiro Yoshikawa received the Ph.D. degree in engineering from Osaka University, Japan, in 2005.

From April 2003 to March 2005, he was a Research Fellow of the Japan Society for the Promotion of Science (JSPS fellow, DC2). From April 2005 to March 2006, he was a Researcher at Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International. Since April 2006, he has been a Researcher at Asada Synergistic Intelligence Project, ERATO, Japan Science and Technology Agency. He has been engaged in the issues of humanrobot interaction and cognitive developmental robotics.

Masaki Ogino received the B.S., M.S., and Ph.D. degrees from Osaka University, Osaka, Japan, in 1996, 1998, and 2005, respectively.

He was a Research Associate in Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University from 2002 to 2006. He is currently a Researcher in Asada Synergistic Intelligence Project of ERATO (Exploratory Research for Advanced Technology by Japan Science and Technology Agency). His research interests are humanoid robot control, biped walking and cognitive issues involved in humanoid robots.

Chisato Yoshida received the M.A. and Ph.D. degrees in experimental and cognitive psychology from Kobe University, Kobe, Japan.

Since 2000, she has served as a Postdoctoral Researcher at Graduate School of Informatics, Kyoto University, to research human spatial cognition, visuo-motor transformation, and their brain mechanisms. She engaged in researches on human perceptual properties and mechanisms of gaze and eye contact at ATR Human Information Science Laboratories since 2005. Since 2007, she has been a Researcher of Japan Science and Technology Agency for ERATO Asada Synergistic Intelligence Project, and going on psychological researches about functional development and neural mechanisms of motor control to differentiate social cognition during human infancy and childhood.