Teaching a Humanoid Robot through Physical Feedback: So Easy Even a Five Year Old Could Use It

Dai HASEGAWA Rafal RZEPKA Kenji ARAKI
Hokkaido University
{hasegawadai, kabura, araki}@media.eng.hokudai.ac.jp

Abstract: We propose a communication robot that can acquire verb meanings through interaction with a human. Our target verbs are body movement verbs that have two objectives and can be performed only by using arms and head of a humanoid robot. In our system, sentences as "place the right hand on the head", are input using a keyboard and the movements are taught from a human by moving the robot's arms through physical interaction. For an impression evaluation using the SD method, 16 participants taught the robot two verbs, which are "put-on" and "move-away-from", through ten times interactions with the robot. The result showed the verb acquisition ability using direct physical feedback is efficient to make human-robot communication more interesting, enjoyable and fulfilled.

1 INTRODUCTION

By the year 2025, the world's problem with aging society will become painful. Around 1.2 billion people are age 60 and over in the world and this number is expected to rise to 2 billion by 2050, according to WHO's estimates[1]. Moreover, many elderly people will live alone and people will have to spend much more time for work, because work force decreases. In such society, one of the most important problems is decreasing chances of communication among people. Therefore, developing robots that can communicate with human is needed to improve the quality of our life.

There are already several robots defined as "communication robots" that their purpose is to communicate with humans [1, 2, 3], and they are able to communicate with human through many types of media [4, 5, 6], which are for example speech-language, body movements or facial expressions. However, the quality of communication is too low for users to use them for a long time because they achieved only simple and shallow communication, for example, interactions based on the static rules and conversations using limited vocabulary. To make human-robot communication more satisfying, we strongly believe that more complex and deeper communication is needed, and language acquisition is one of abilities to realize such communication through teaching a machine word meanings.

In this paper, we will propose novel language acquisition method through physical feedback in a humanoid robot to communicate with humans. Physical interaction is important for enjoying communication, and verbs are most basic words to be taught by physical feedback.

We will describe related works and original points of our language acquisition method in the section 2, and explain our language acquisition system in the following sections. Then we will give a brief explanation about a verb acquisiton experiment that shows the robot can learn four actions: "place-on", "move-close-to", "move-away-from", and "touch-with" through interaction with human, and the learned verbs have robustness in terms of changing objectives, initial position, and end position in the section 8, and in the section 9 we will describe how the verb teaching interaction influences users through a simple human-robot interaction experiment. In these experiments our target language is Japanese and we will use *italic* when giving Japanese examples. Because our method is language independent, we will examine it with other languages as English, in the future. Lastly, we will give discussions and our conclusions.

[1]WHO project, Ageing and life course: http://www.who.int/ageing/en/

2 STATE OF THE ART

To make robots acquire language, we have to face the symbol grounding problem [7] how computers automatically relate the symbolic language system to the non-symbolic real world. Several ways of symbol grounding method are proposed [8, 9, 10, 11, 12, 13, 14]. However, we believe that there are still many problems that should be resolved in verbs acquisition, although many researchers have tried to realize the system. This is because a motor pattern that robots have to generate completely changes corresponding to both language contexts (objectives) and physical contexts (initial position, end position, obstacles) which are input a verb.

There are some research activities in verb acquisition field [15, 16]. Tani et al. [17] described a system using Recurrent Neural Network, where a movable arm robot acquires nouns and verbs from pairs of a two words phrase like "push green" and a motor pattern. Sugiura et al. [18] also developed a verbs acquisition model for an arm robot. They used Hidden Markov Model to learn object-manipulation-verb meanings from sets of a sentence and a trajectory of robot's arm. The trajectory was in the trajector[2]-reference point[3] specific coordinate system.

However, their verb representation models are statistical models based on direct motor patterns or the trajector's trajectories. The model based only on trajectories

[2]A trajector is what moves mainly in a movement. It can be an object or a body part.

[3]A reference point is what is referred by a trajector in a movement.

can not represent meanings of some verbs which are independent of the trajectories. For instance, "move the right hand close to the left hand" does not mean the way how the right hand moves to the left hand but how the distance between the right hand and the left hand was shorten independently of its trajectory. Moreover, though they achieve visionary teaching methods to teach movements to robots, we believe that physical contact teaching is also efficient for more entertaining communication.

Original points of our proposed method are a movement teaching method using direct physical feedback and a feature based verb representation model based not only on trajectories but also on trajector-reference point relationships. We will describe a feature based representation model that has six features including both trajector-reference point relationships and the trajector's trajectory for body manipulation movements. Body movement verbs, our targets, are verbs which imply body movement and which have two objectives which also imply body parts. Our representation model has merits of the fact that some verbs which are independent of the trajectories are represented appropriately and it is easy for a designer to understand how these verbs are represented. Above mentioned probabilistic and connectionist methods do not have these merits. We also describe a novel learning algorithm where a humanoid robot acquired abstract verb meanings from sets of a textual command and a motor pattern which are taught by human using direct physical feedback.

3 OVERVIEW OF OUR SYSTEM

We will show an overview of our system below (Figure 1). It works in two phases: learning and testing. In the learning phase, a user inputs a Japanese textual command which contains a body movement verb with two objects by using a keyboard. Then, the user also inputs a proper movement to the robot through direct physical feedback (see section 3.3). The feedback movements are detected as motor angle patterns retrieved from each sensor. Next, the system converts a set of a command and a motor pattern to a set of the command and an movement representation in Movement Cognition Module, and then adds it to the Example Database. From actual and concrete examples of movement representation of a verb in the Example Database, the system creates a meaning of the verb by abstracting the examples and adds it to the Rule Database in Abstraction Module.

In the testing phase, a user inputs an unknown sentence which contains an already known verb in unknown language contexts (objectives) and physical contexts (initial position, end position, obstacles). Then, the system generates a motor angle pattern using a rule and outputs an movement.

3.1 Prerequisite

In this verbs grounding algorithm, we assume that the system already acquired Japanese morphology, because in Japanese text processing the computer has to segment a sentence into morphological elements first as there are no spaces between words. In our system we use a morphological analyzer MeCab[4] to segment sentences. Moreover, we also assume that the system has acquired nouns about the robot's body parts. That is, the system can understand which part of body is "the right hand" and can calculate where "the right hand" is exactly in the standard coordinate system (see section 3.2).

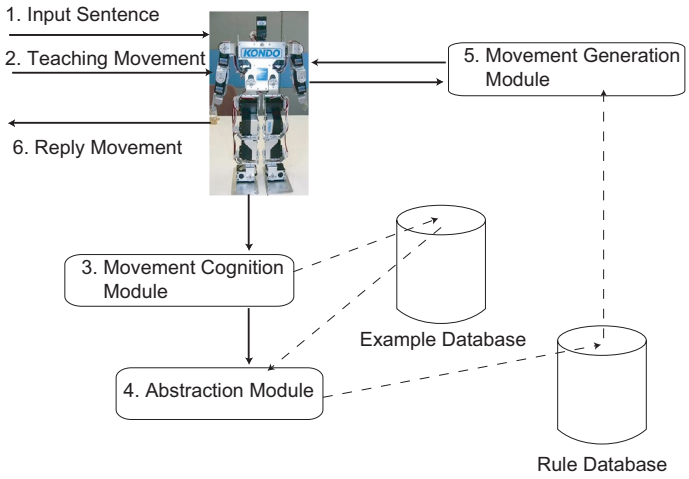

Fig. 1: System Overview

3.2 HUMANOID ROBOT

For our experiments, we used a humanoid robot (KHR2-HV[5]) shown in Figure 2. The robot is equipped with 17 motors but no sensors, and it sends signals describing only its motor states.

We set the standard coordinate system where the origin is at the robot's chest, the x-axis is the horizontal direction of robot's front side, and the z-axis is vertical (see Figure 2).

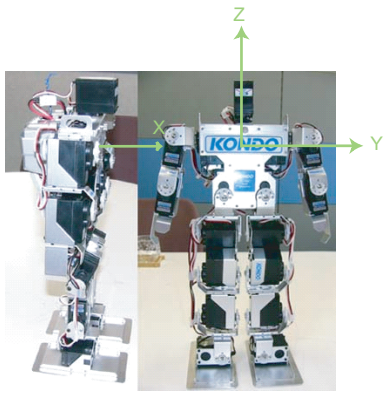

Fig. 2: KHR2-HV

[4]MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.jp/

[5]Kondo Kagaku Co. Ltd, http://www.kondo-robot.com/

3.3 DIRECT PHYSICAL FEEDBACK

Several methods have already been developed where a human supervisor teaches movements to humanoid robots, e.g. the vision based method [19, 20] or the motion capture based method [21]. However, we decided to implement a direct physical feedback method where humans teach movements to a robot by actually moving its body parts. We claim that it is a universal and natural method which allows teaching new movements within the limits of any humanoid robot's body structure. In addition it needs no extra equipment as cameras and microphones and can be implemented in even very simple and inexpensive toy humanoids.

4 REPRESENTATION MODEL

The below is the proposed representation model for body movements. In this model, we define six features to represent movements. Our target verbs depend on the robot's structure we use (e.g. the robot cannot grab things), and we set the six features to represent these verbs. The system can represent movements which are independent of the trajector's trajectory, because the 3rd, 4th and 5th features describe relationship between trajector and reference point. The system automatically generates the representation from a sentence and a motor pattern. Figure 3 is an example of "place the right hand on the head."

1. A trajector name
2. A reference point name
3. The variation of distance between trajector and reference from initial to final position
4. The distance between trajector and reference in the final position
5. An above/under positional relationship of trajector to object
6. A trajector's trajectory in the coordinate system where the origin is at reference point's position, x-axis is the horizontal direction of trajector's initial position, and z-axis is vertical.

place the right hand on the head

```
1: "the right hand"
2: "the head"
3: -10.486
4: 2.185
5: "above"
6: 9.83 -0.09 -2.34
   10.1 -0.21 -2.35
   10.0 -0.39 -1.88
   10.4 -2.17 0.32
   10.2 -3.43 1.55
   8.02 -3.47 2.92
   6.41 -2.64 3.06
   ...
```

Fig. 3: Example of the Model

5 MOVEMENT COGNITION MODULE

A set of a textual command and a motor angle pattern which is input by a user is transformed to an example with the representation model. We will explain how the Movement Cognition Module works, considering an example "put the right hand on the head" (Figure 4). First, the motor angle pattern is converted to trajectories of referenced body parts, "the right hand" and "the head", by solving the direct kinematics problem. Then, the system distinguishes which noun is a trajector and which one is a reference point. Next, the system calculates values of other features. Finally, the system adds the set of a command and an movement representation to the Example Database.

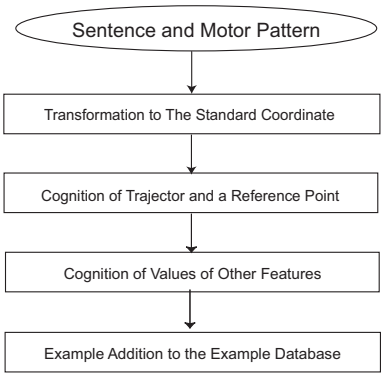

Fig. 4: Movement Cognition

6 ABSTRACTION MODULE

In the Abstraction Module, all examples (which are sets of a command and an movement representation) about one verb are abstracted as one rule. We will describe details of the process here (Figure 5). First, all examples' strings of nouns in language part are parameterized as "@1" and "@2", and corresponding first and second features in the movement representation part are parameterized as the same strings. In the following process, all examples which have both the same abstract sentence and the same abstract first and second features are regarded as targets of verb abstraction. Then, the system determines feature importance from the 3rd, the 4th and the 5th features by comparing all examples about each verb. Next, values of determined features are averaged as values of rule's features. Finally, the system saves all rules to the Rule Database.

7 MOVEMENT GENERATION MODULE

In the testing phase, a textual command input by a user is processed in the Movement Generation Module shown in Figure 6. First, the system distinguishes a trajector and a reference point in a command which contains known verb corresponding to the rule about the verb. Then, the system determines the final position of an movement with the 3rd, the 4th, and the 5th feature of the rule. Next, the trajectory of overall movement is generated, suiting the final position. Finally, the trajectory is translated to a motor angle pattern. Below we will explain how the system distinguishes a trajector and a reference point, and how it creates the trajectory.

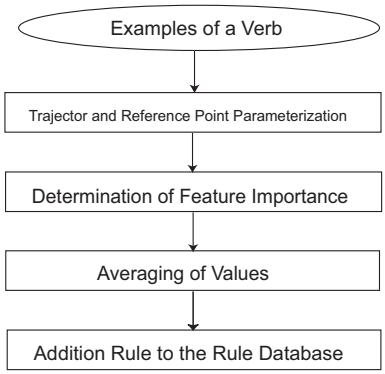

Fig. 5: Abstraction Module

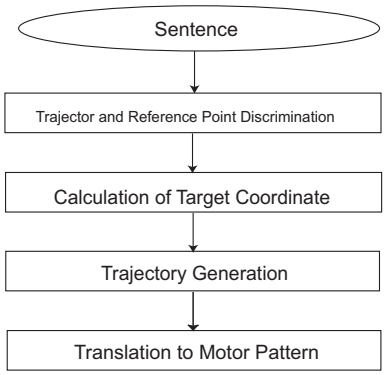

Fig. 6: Movement Generation Module

8 VERB ACQUISITION EXPERIMENT

We implemented the above mentioned algorithm in the humanoid robot, and conducted body movement verbs acquisition to confirm if the robot can acquire verbs which have robustness in terms of combination of objectives, initial position, end position and an obstacle. We set target verbs as "*oku* (place-on)", "*chikazukeru* (move-close-to)", "*hanasu* (move-away-from)", "*sawaru* (touch-with)."

Our experimental design is described below. First, for each verb, the robot learns two training sentences with different combination of objectives, e.g. "place the right hand on the head" and "place the left hand on the right hand", three times each in different initial and end position. Then, a test sentence with unknown combination of objectives is tested three times in different initial and end position. We set training combinations of objectives as "the right hand and the head" and "the left hand and the right hand." Then we also set the test combination as "the left hand and the head." If the robot outputs proper movements, we regard the verb as acquired.

Two participants conducted both learning task and test. Then they evaluated the output movements of a test sentence three times on a three point scale (3 is proper, 1 is wrong). Table 1 shows the experimental results.

Table 1: Results (3 is proper, 1 is wrong)

Verb	Ave. of A	Ave. of B	Ave.
place-on	2	2.4	2.2
move-close-to	2	2.7	2.4
move-away-from	3	2	2.5
touch-with	3	2.4	2.7

9 IMPRESSION EVALUATION

To find out how human feels about teaching communication, we conducted an impression evaluation using human-robot interaction. In this experiment we had sixteen participants, which were seven males aged 20-30, two males aged 30-40, four females aged 20-30 and three females aged 30-40. Figure 7 shows settings of this experiment. They were asked to have three types of interaction with the humanoid robot which had a body and a GUI (Figure 8). The types of interaction are described as below.

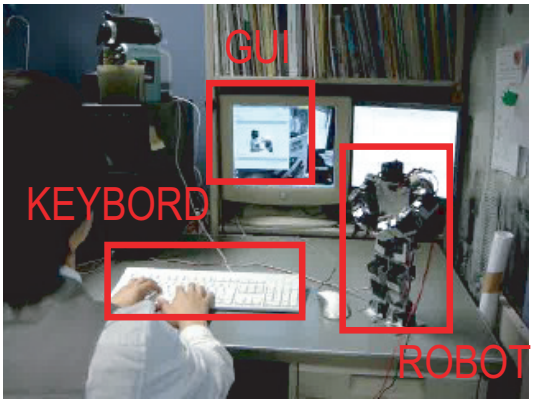

Fig. 7: General Setting

Fig. 8: Graphical User Interface of Our System

1. Teaching communication through physical interactions (System A). The robot does not have any verb

Table 2: Results of Impression Evaluation

	System A	System B	System C
Average	5.34	4.81	4.47

knowledge in the beginning and learns verb meanings.

2. Order-Reply interaction (System B). The robot knows verbs meanings, and outputs a movement for all orders.

3. Teaching communication through physical interactions (System C). The robot does not have any verb knowledge in the beginning and does not learn verb meanings.

In all types of communication, participants input utterances using keyboard and the robot replied with textual utterances or movements, and all systems knew noun meanings in the beginning. The users made utterances using one verb and two objective nouns, for example "*migite wo atama ni oite* (place the right hand on the head)." The verbs were chosen from "*oku* (place-on)" and "*hanasu* (move-away-from)," while the nouns were chosen from "*migite* (the right hand)," "*hidarite* (the left hand)" and "*atama* (the head)."

The procedures of each communication are described as below. In system A, when a user inputs an utterance and the robot does not know the verb meaning, it asks "*migite wo atama ni oite wo oshiete* (teach me to place the right hand on the head)." Then the user teaches a movement. Whereas, if the robot knows the verb meaning, it outputs a movement that learned before, and the user evaluates if the movement is correct. If it is, the user pushes a "OK" button, and if it is wrong the user performs teaching again. In system B, when a user inputs an utterance, the robot outputs a correct movement using initial verb knowledge. In system C, when a user inputs an utterance the robot can not output any movement and asks the user to teach it. Then the user teaches a movement.

We explained the systems' features for participants in advance, and they had 10 times interactions according to above procedures per each system in random order. Shortly after the interactions participants evaluated systems they used. To measure impressions of the systems, we used Semantic Differential method with 30 adjective pairs. Participants were asked to evaluate the interactions using seven levels of all given pairs and the results are shown in Table 2 and Figure 9.

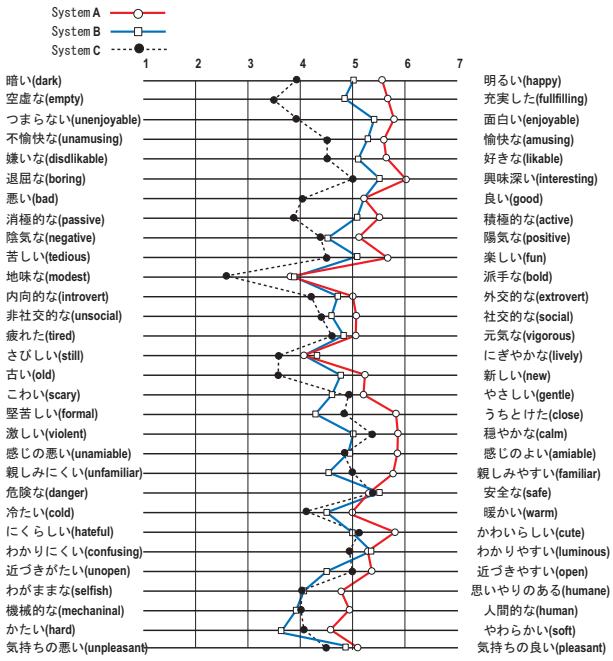

Fig. 9: Impressions

consider the capacity and limitations of the model than the statistical representation models.

The impression evaluation showed that system A had better impression evaluation as a communication robot than others. Therefore, we can conclude that the teaching interaction, which is one of complex and deep comunication, is more enjoyable and more interesting for human than simple and shallow order-reply interaction. Moreover, we found that if teaching is finally not accomplished, the impression evaluation values decrease..

```
place @A on @B

1: "@A"
2: "@B"
3: -11.447
4: 2.175
5: "above"
6: 8.43, -0.01, -6.17
   8.47, -0.03, -6.15
   8.53, -0.09, -6.16
   8.40, -0.18, -6.06
   8.32, -1.37, -3.26
   7.13, -1.80, -0.30
   5.26, -1.25, 1.561
   ...
```

Fig. 10: place-on

10 DISCUSSIONS

The verb acquisition experiment showed that the robot properly acquired four problematic verbs, "place-on", "move-close-to", "move-away-from", and "touch-with" by being taught the movements only six times in different contexts. Figures 10 and 11 show how the system acquired the meanings of the verbs. These representations clearly show what the important features are for verb acquisition for robots. That is an important part of a robot design process. Using our method makes it easier to understand how the robot represents verbs and to

11 CONCLUSIONS

We proposed an algorithm where a humanoid robot acquires body manipulation verbs through more complex and deeper communication with human. The algorithm includes: physical feedback, a novel representation model with six features containing both the trajector-reference point relationships and trajector's trajectory for movements, a mechanism that creates the abstract verb meanings from sets of a textual command and a movement

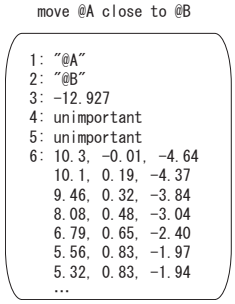

Fig. 11: move-close-to

representation, and a process that generates movements for unknown inputs. As a result of the verb acquisition experiment, the humanoid robot properly acquired four problematic verbs "*oku* (place-on)", "*chikazukeru* (move-close-to)", "*hanasu* (move-away-from)", and "*sawaru* (touch-with)" where the abstract meanings of verbs were independent of contexts. Furthermore, the impression evaluation results indicated that the more complex and deeper the communication is the more interasting, enjoyable, and fulfilled human-robot communication becomes. In the future work, we would like to extend language acquisition ability.

References

[1] S. Kato, S. Ohshiro, H. Itoh, and K. Kimura, "Development of a communication robot ifbot," in *In Proceedings of the 2004 IEEE International Workshop on Robotics and Automation*, pp. 697–702, April 2004.

[2] F. Tanaka and H. Suzuki, "Dance interaction with qrio: a case study for non-boring interaction by using an entrainment ensemble model," in *In Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication*, pp. 419–424, September 2004.

[3] C. G. Burgar, P. S. Lum, P. C. Shor, and H. M. V. der Loos, "Development of robots for rehabilitation therapy : The palo alto va/stanford experience," *Journal of Rehabilitation Research and Development*, vol. 37, pp. 663–673, November/December 2000.

[4] J. Cassell, M. Steedman, N. Badler, C. Pelachaud, M. Stone, B. Douville, S. Prevost, and B. Achorn, "Modeling the interaction between speech and gesture," in *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 153–158, 1994.

[5] J. Cassell, H. Vilhjalmsson, and T. Bickmore, "Beat: the behavior expression animation toolkit," in *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 477–486, 2001.

[6] D. Hasegawa, J. Sjoberghand, R. Rzepka, and K. Araki, "Automatically choosing appropriate gestures for jokes," in *To appear the Proceedings of the Fifth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-09)*, (San Francisco, USA), October 2009.

[7] S. Harnad, "The symbol grounding problem," *Physica D*, vol. 42, pp. 335–346, 1990.

[8] D. Roy, "Learning words from sights and sounds: A computational model," *Cognitive Science: A Multidisciplinary Journal*, vol. 26, pp. 335–346, January 2002.

[9] D. Roy, *A Mechanistic Model of Three Facets of Meaning. Chapter to appear in Symbols, Embodiment, and Meaning, de Vega, Glenberg, and Graesser, eds.* Oxford: Oxford Press, 2008.

[10] N. Iwahashi, "Language acquisition by robots? towards a new paradigm of language processing," *Journal of Japanese Society for Artificial Intelligence, Special Issue on Language Acquisition*, vol. 48, no. 1, pp. 49–58, 2003.

[11] L. Steels and F. Kaplan, "Aibo's first words, the social learning of language and meaning," *Evolution of Communication*, vol. 4, no. 1, pp. 3–32, 2001.

[12] L. Steels, "Language games for autonomous robots," *IEEE Intelligent Systems*, vol. Sep/Oct issue, pp. 16–22, 2001.

[13] L. Steels, "Semiotic dynamics for embodied agents," *IEEE Intelligent Systems*, vol. 21, pp. 32–38, 2006.

[14] D. Hasegawa, R. Rzepka, and K. Araki, *Connectives Acquisition in a Humanoid Robot Based on an Inductive Learning Language Acquisition Model*. Vienna, Austria: To appear in Humanoid Robots, I-Tech Education and Publishing, 2008.

[15] J. M. Siskind, "Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic," *Journal of Artificial Intelligence Research*, vol. 15, pp. 31–90, 2001.

[16] R. A. Peters and C. L. Campbell, "Robonaut task learning through teleoperation," in *In Proceedings of the 2003 IEEE International Conference on Robotics and Automation*, (Taipei, Taiwan), pp. 23–27, September 2003.

[17] Y. Sugita and J. Tani, "Learning semantic combinatoriality from the interaction between linguistic and behavioral processes," *Adaptive Behavior*, vol. 13, no. 1, pp. 33–52, 2005.

[18] K. Sugiura and N. Iwahashi, "Motion recognition and generation by combining reference-point-dependent probabilistic models," in *Proceedings of IEEE/RSJ 2008 International Conference on Intelligent Robots and Systems (IROS 2008)*, (Nice, France), pp. 852–857, September 2008.

[19] M. J. Mataric, "Getting humanoids to move and imitate," *IEEE Intelligent Systems*, vol. 15, pp. 18–24, July–August 2000.

[20] S. Schaal, "Is imitation learning the route to humanoid robots?," *Trends in Cognitive Science*, vol. 3, pp. 233–242, June 1999.

[21] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, "Embodied symbol emergence based on mimesis theory," *The International Journal of Robotics Research*, vol. 23, no. 45, pp. 363–377, 2004.