A Method for Acquiring Body Movement Verbs for a Humanoid Robot through Physical Interaction with Humans

Dai Hasegawa, Rafal Rzepka, and Kenji Araki
Graduate School of Information Science and Technology
Hokkaido University
Japan
{hasegawadai, kabura, araki}@media.eng.hokudai.ac.jp

Abstract

In this paper, we propose a language grounding method that relates verbs which imply body manipulation to motor angle patterns in a humanoid robot to make robots more entertaining. In established methods, verbs are represented by statistical models based on trajectories or motor patterns of a trajector. In our method we use a novel representation model that has six features including both trajector-reference point relationships and the trajector's trajectory. By using this model, some verbs which do not depend on a trajectory, e.g. "move the right hand close to the left hand." are represented more adequately. In our language grounding method a humanoid robot generates abstract verb meanings independent of context. As input it uses sets of a user input textual command and a motor pattern. The motor pattern is taught using direct physical feedback resembling playing with child. We implemented the algorithm in a humanoid robot and conducted a verb acquisition experiment. As a result, four problematic verbs, "place-on", "move-close-to", "move-away-from", and "touch-with" were acquired correctly.

Introduction

Recently, many robots are developed to amuse people at home (Kato et al. 2004; Tanaka and Suzuki 2004), not to perform tasks in a factory. Some of such robots are able to have a conversation with human and to output movements. Many people use them to relax, have fun, or perform rehabilitation (Burgar et al. 2000). However, the robots have a problem to entertain us for longer time. That is, they always talk using limited vocabulary, reply patterns, and kinds of movement. For this reason, users soon become bored. Therefore, we strongly believe that robots having an ability to learn language and movements through interaction with human make people happier. That ability is the same as the language acquisition ability. However, as far as acquiring language by robots, we have to face the symbol grounding problem (Harnad 1990) how the computer automatically relates the symbolic language system to the non-symbolic real world.

There are several research activities on language acquisition of embodied systems and several ways of symbol grounding are proposed. Nouns and adjectives grounding have been established by connecting object names to perceptual categories in visual and audio perception of an arm robot using a statistical cross-modal model (Roy 2002; 2008) or the Hidden Markov Model (Iwahashi 2003). Steels developed a language grounding model through language game (Steels and Kaplan 2001; Steels 2006).

Whereas nouns and adjectives acquisition have been successful, we believe that there are still many problems which should be resolved in verbs acquisition from the research focused on connectives acquisition in a humanoid robot (Hasegawa, Rzepka, and Araki 2009), 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 context (objectives) and physical context (initial position, end position, obstacles) which are input a verb.

State of the Art

We believe that verbs acquisition difficulties in a computer are derived from higher ambiguity than nouns'. When a robot makes a movement for "place-on", the motor patterns which should be generated will change significantly depending on contexts such as an initial motor pattern, a position of objects referred by objectives, an existence or nonexistence of obstacles, and so on. That is why the robot needs a representation which is able to describe meanings of verbs independent of context.

There are some research activities in verb acquisition (Siskind 2001; Peters and Campbell 2003). Tani et al. (Sugita and Tani 2005) 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. (Sugiura and Iwahashi 2008) 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[1]-reference point[2] specific coordinate system.

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

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

However, their verb representation models are statistical models based on direct motor patterns or the trajector's trajectories. We propose a novel representation model for movements based not only on trajectories but also trajector-reference point relationships, because the model based only on trajectories 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.

Original points of our method proposed in this paper are a feature based representation model for verbs and a movement teaching method using direct physical feedback. 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 that some verbs which are independent of the trajectories are represented appropriately and it is easy for designer to understand how verbs are represented. Above mentioned probabilistic and connectionist method do not have these merits. We also show an 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. At last, we give a brief explanation about experimental results where 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, end position, and obstacles. In this experiment, 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 language, e.g. English, in the future.

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 of **Direct Physical Feedback**). 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 about 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 context (objectives) and physical context (initial position, end position, and obstacles). Then, the system generates a motor angle pattern using a rule and outputs an movement.

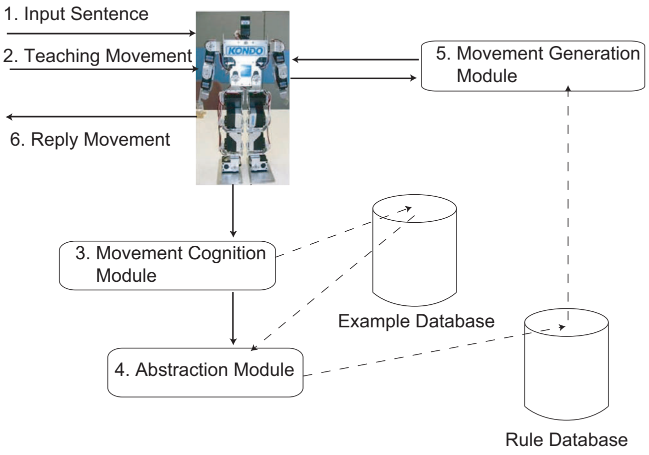


Figure 1: System Overview

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. In our system we use a morphological analyzer MeCab[3] 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 of **Humanoid Robot**).

Humanoid Robot

For our experiments, we used a humanoid robot (KHR2-HV[4]) 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).

Direct Physical Feedback

Several methods have already been developed where a human supervisor teaches movements to humanoid robots, e.g. the vision based method (Mataric 2000; Schaal 1999) or the motion capture based method (Inamura et al. 2004). 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.

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

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

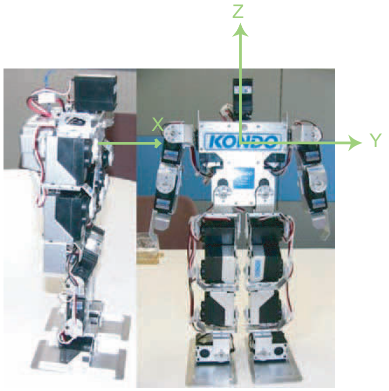


Figure 2: KHR2-HV

In addition it needs no extra equipment as cameras and microphones and can be implemented in even very simple and inexpensive toy humanoids.

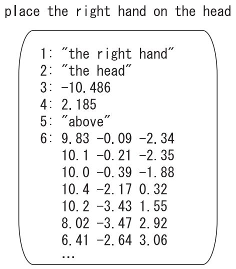
Representation Model

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 can not 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.

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.



```
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
   ...
```

Figure 3: Example of the Model

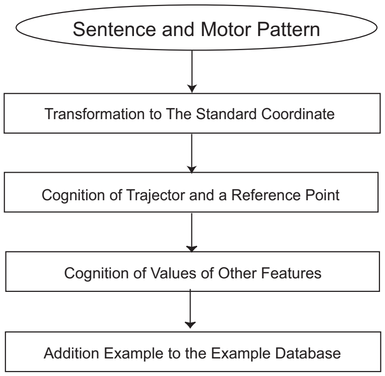

Sentence and Motor Pattern
Transformation to The Standard Coordinate
Cognition of Trajector and a Reference Point
Cognition of Values of Other Features
Addition Example to the Example Database

Figure 4: Movement Cognition

Cognition of Trajector and Reference Point

Cognition of a trajector and a reference point are based on an assumption that a trajector moves the largest amount of distance in all candidates. Thus, the system can distinct a trajector as the one which has larger moving distance than the other.

Cognition of Values of Other Features

The values of other features are calculated after cognition of a trajector and a reference point. Where $\mathbf{P_0}$ is the initial position of trajector in the standard coordinate system, $\mathbf{P_L}$ is the final position of trajector, $\mathbf{O_0}$ is the initial position of reference point, $\mathbf{O_L}$ is the final position of object. Then, the 3rd feature is calculated (1), the 4th feature (2), and the 5th feature (3). The 6th feature is given by solving the forward kinematics problem about $\mathbf{P_i}$ $(i = 0...L)$.

$$3rd\ value = |\mathbf{P_0} - \mathbf{O_0}| - |\mathbf{P_L} - \mathbf{O_L}| \qquad (1)$$

$$4th\ value = |\mathbf{P_L} - \mathbf{O_L}| \qquad (2)$$

$$5th\ value = \begin{cases} ''above'' & \text{if } \mathbf{P_{Lz}} >= \mathbf{O_{Lz}}, \\ ''under'' & \text{if } \mathbf{P_{Lz}} < \mathbf{O_{Lz}} \end{cases} \qquad (3)$$

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 feature in the movement representation part are parameterized as the same strings. In the process from here, all examples which have both the same abstract sentence and the same abstract first and second feature are regarded as targets of abstraction of a verb. Then, the system determines feature importance from the 3rd, the 4th and the 5th 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.

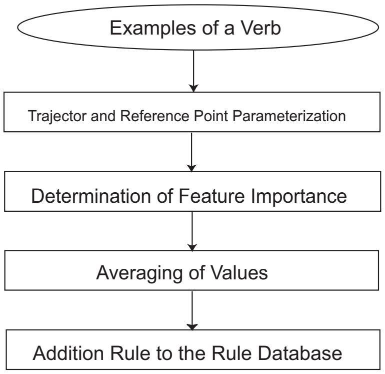

Examples of a Verb
Trajector and Reference Point Parameterization
Determination of Feature Importance
Averaging of Values
Addition Rule to the Rule Database

Figure 5: Abstraction Module

Determination of Important Features

For the process of a feature importance determination, we assume that features which have high similarity among all examples are important. That is, despite the examples which given in various contexts, having similar values means that the features are the verb specific features to represent the verb beyond contexts. Thus, we define similarity for the 3rd, the 4th, and the 5th features using (4), (5), and (6).

$$the\ 3rd\ feature = \begin{cases} important & \text{if } x < 0.1, \\ important & \text{if } x > 0.9, \\ unimportant & \text{oterwise.} \end{cases} \tag{4}$$

$$where\ x = \frac{the\ number\ of\ positive\ samples}{the\ number\ of\ samples}$$

$$the\ 4th\ feature = \begin{cases} important & \text{if } \sigma^2 < 2.0, \\ unimportant & \text{oterwise.} \end{cases} \tag{5}$$

$where\ \sigma^2{=}\frac{1}{N}\sum_{k=1}^{N}(\overline{f_4}{-}f_{4k})^2,\ N{:}the\ number\ of\ samples,$

$\overline{f_4}{:}the\ average\ of\ values,\ f_{4k}{:}value\ of\ sample\ k$

$$the\ 5th\ feature = \begin{cases} important & \text{if } x < 0.1, \\ important & \text{if } x > 0.9, \\ unimportant & \text{oterwise.} \end{cases} \tag{6}$$

$$where\ x = \frac{the\ number\ of\ above\ samples}{the\ number\ of\ samples}$$

In these equations, we determined the thresholds in a preliminary experiment.

Trajectories Averaging

The trajectories averaging process also includes a process of minimizing them. Trajectories of samples have various numbers of plots. However, we believe that the trajectory has to influence movements as slightly as possible. For this reason, the system averages and minimizes the trajectories according to (7) and (8).

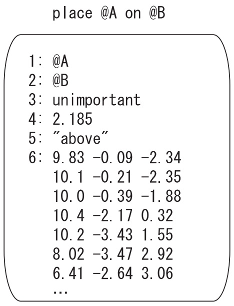
$$\mathbf{P_r} = \{\mathbf{P_{r1}}, \mathbf{P_{r2}}, \mathbf{P_{r3}}, ..., \mathbf{P_{rL}}\} \tag{7}$$

$$\mathbf{P_{ri}} = \frac{1}{N}(\mathbf{P_{1i}} + \mathbf{P_{2i}} + \mathbf{P_{3i}} + ... + \mathbf{P_{Ni}}) \tag{8}$$

$where\ N{:}the\ number\ of\ samples,\ L{:}the\ number\ of\ minimum\ plots,$

$\mathbf{P_{ri}}{=}(x{,}y{,}z){:}\ the\ plot\ i\ of\ rule,\ \mathbf{P_{ki}}{=}(x{,}y{,}z){:}\ the\ plot\ i\ of\ example\ k$

We will show an example of rule as Figure 6.



```
place @A on @B

1: @A
2: @B
3: unimportant
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
   ...
```

Figure 6: Example of Rule

Movement Generation Module

In the testing phase, a textual command input by a user is processed in the Movement Generation Module shown as in Figure 7. 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.

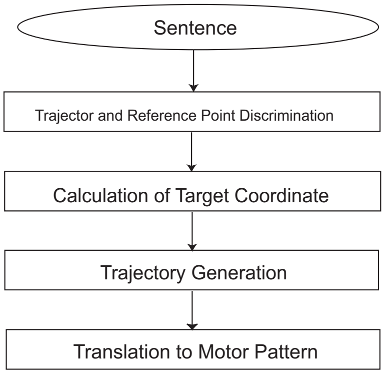

Sentence
Trajector and Reference Point Discrimination
Calculation of Target Coordinate
Trajectory Generation
Translation to Motor Pattern

Figure 7: Movement Generation Module

Discrimination of Trajector and Reference Point

The discrimination of a trajector and a reference point is preceded as below. First, nouns of the input command are replaced by ”@1” and ”@2” in order of appearance. Then, the system extracts a rule which has the same abstract command, and the 1st and 2nd feature values of rule indicate which is a trajector and which a reference point between ”@1” and ”@2”, that is, the system can find out which noun is a trajector and the other is a reference point.

Calculation of Target Coordinate

After the distinction of a trajector and a reference point, the system calculates a target coordinate of movement with the 3rd, the 4th, and the 5th feature of the rule. The target coordinate is the last position of movement. In this process, the system uses only the features which are determined as important features for the verb. However, the 3rd and the 4th feature are incompatible in the determination of the target coordinate. Thus, we have to set two features priorities. For this research, we decided that the 4th feature is prior to the 3rd one, though, we believe that this priority is not important for generating movements. It is just a temporal treatment. This is because if the proper teaching was enough then only one feature would remain as an important feature between the 3rd and the 4th feature. For this reason, the system calculates the target coordinate with the 3rd or the 4th and the 5th feature. Moreover, we introduce another constraint in this process. That is the minimum cost constraint where the robot always selects the nearest point for an movement final position from the candidates.

Trajectory Generation

The system generates a movement trajectory corresponding to the decided target coordinate. However, the target coordinate is represented in the standard coordinate system but a rule’s 6th feature, the averaged trajectory, is represented in the trajector-reference point specific coordinate system. Thus, the system, in advance, transforms the trajectory from the trajector-reference point specific coordinate system to the standard coordinate system after creating the trajector-reference point specific coordinate system which depends on current positions of both a trajector and a reference point. Finally, the system also calculates a parallel transformation of the trajectory to suit its final position to the target coordinate.

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

Verb	Ave. of A	Ave. of B	Ave.
place-on	2	1.6	1.8
move-close-to	2	1.3	1.6
move-away-from	1	2	1.5
touch-with	1	1.6	1.3
stroke	2.3	2.6	2.5
peck-with	2.3	3	2.6

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)”, ”*naderu* (stroke-with)”, ”*kosuru* (peck-with).”

Our experimental design is described below. First, for each verb, the robot learn 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 output proper movements, we regard the verb is acquired.

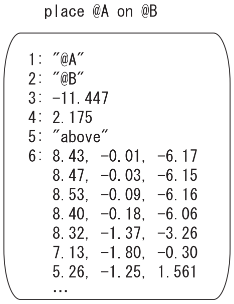
Two participants, which are a male and a female aged 20-30, conducted both learning task and test. Then they evaluated the output movements of a test sentence three time on a three point scale (1 is proper, 3 is wrong). Table 1 shows the experimental results.

Discussion

The 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 8 and 9 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 consider the capacity and limitations of the model than the statistical representation models.

Moreover, our representation has another merit. That is, the robot can apply the acquired verbs knowledge to situations where an obstacle in its way. We simulated an obstacle and made the robot output movement using a simple program that the robot manipulates its arm around obstacle unless important features change. As the result, the robot properly output movements for the acquired four verbs. Figure 10 shows the internal representation about ”put-on” in this simulation.

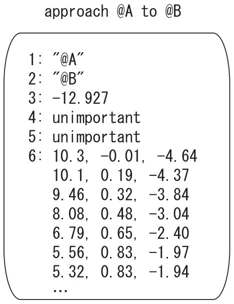
While, the representation model could not represent ”stroke-with” and ”peck-with” that the movement process is important. This is because our representation does simple averaging for trajectories. We need an abstraction algorithm for trajectories in the future, and we also have a plan to evaluate efficienciy of our method in other robots.



```
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
   ...
```

Figure 8: place-on



```
approach @A to @B

1: ″@A″
2: ″@B″
3: -12.927
4: unimportant
5: unimportant
6: 10.3, -0.01, -4.64
   10.1, 0.19, -4.37
   9.46, 0.32, -3.84
   8.08, 0.48, -3.04
   6.79, 0.65, -2.40
   5.56, 0.83, -1.97
   5.32, 0.83, -1.94
   ...
```

Figure 9: move-close-to

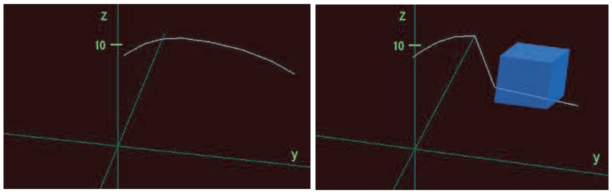


Figure 10: trajectory of ”place-on” with obstacle

Conclusion

We proposed an algorithm where a humanoid robot acquires body manipulation verbs for entertaining users with its learning capability. The algorithm includes: 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 representation, and a process that generates movements for unknown inputs. As a result of our experiment, the humanoid robot properly acquired four problematic verbs ”*oku* (place-on)”, ”*chikazukeru* (move-close-to)”, ”*hanasu* (move-away-from)”, and ”*sawaru* (touch-with)”and the abstract meanings of verbs were independent of contexts. In the future work, we would like to investigate how efficient our language acquisition mechanism is in entertaining.

References

Burgar, C. G.; Lum, P. S.; Shor, P. C.; and der Loos, H. M. V. 2000. Development of robots for rehabilitation therapy : The palo alto va/stanford experience. *Journal of Rehabilitation Research and Development* 37(6):663–673.

Harnad, S. 1990. The symbol grounding problem. *Physica D* 42:335–346.

Hasegawa, D.; Rzepka, R.; and Araki, K. 2009. *Connectives Acquisition in a Humanoid Robot Based on an Inductive Learning Language Acquisition Model*. Vienna, Austria: Chapter in Humanoid Robots, Edited by Ben Choi, In-Tech.

Inamura, T.; Toshima, I.; Tanie, H.; and Nakamura, Y. 2004. Embodied symbol emergence based on mimesis theory. *The International Journal of Robotics Research* 23(45):363–377.

Iwahashi, N. 2003. Language acquisition by robots? towards a new paradigm of language processing. *Journal of Japanese Society for Artificial Intelligence, Special Issue on Language Acquisition* 48(1):49–58.

Kato, S.; Ohshiro, S.; Itoh, H.; and Kimura, K. 2004. Development of a communication robot ifbot. In *In Proceedings of the 2004 IEEE International Workshop on Robotics and Automation*, 697–702.

Mataric, M. J. 2000. Getting humanoids to move and imitate. *IEEE Intelligent Systems* 15(4):18–24.

Peters, R. A., and Campbell, C. L. 2003. Robonaut task learning through teleoperation. In *In Proceedings of the 2003 IEEE International Conference on Robotics and Automation*, 23–27.

Roy, D. 2002. Learning words from sights and sounds: A computational model. *Cognitive Science: A Multidisciplinary Journal* 26(1):335–346.

Roy, D. 2008. *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.

Schaal, S. 1999. Is imitation learning the route to humanoid robots? *Trends in Cognitive Science* 3(6):233–242.

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

Steels, L., and Kaplan, F. 2001. Aibo's first words, the social learning of language and meaning. *Evolution of Communication* 4(1):3–32.

Steels, L. 2006. Semiotic dynamics for embodied agents. *IEEE Intelligent Systems* 21:32–38.

Sugita, Y., and Tani, J. 2005. Learning semantic combinatoriality from the interaction between linguistic and behavioral processes. *Adaptive Behavior* 13(1):33–52.

Sugiura, K., and Iwahashi, N. 2008. 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)*, 852–857.

Tanaka, F., and Suzuki, H. 2004. 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*, 419–424.