Simultaneous estimation of role and response strategy in human-robot role-reversal imitation learning

Tadahiro Taniguchi * **Hiroto Nakanishi** ** **Naoto Iwahashi** ***

* *Ritsumeikan University, Kusatsu, Shiga, Japan (e-mail: taniguchi@ci.ritsumei.ac.jp)*
** *Ritsumeikan University, Kusatsu, Shiga, Japan (e-mail: nakanishi@ci.ritsumei.ac.jp)*
*** *National Institute of Information and Communications Technology, Kyoto, Japan, (e-mail: naoto.iwahashi@nict.go.jp)*

Abstract: In this paper, we describe a novel imitation learning method which enables an autonomous robot to acquire response strategy and to estimate roles through human-robot real-time interaction. The robot becomes able to respond to human user's social action, e.g. bye-bye and shake hand, correctly. We constructed the learning method based on role reversal imitation which is found in human infants in developmental psychological researches. A probabilistic model is proposed which assumes that delayed reactions are stochastically generated by initiative actions. In an experiment, we show a robot hand became able to exhibit correct reaction and estimate whether another's action is an initiative action or a reaction.

Keywords: Learning algorithms, Machine learning, Probabilistic models, Human-machine interface, Man/machine interaction, Adaptation

1. INTRODUCTION

Our living environment is full of diversified physical dynamics which is difficult for an autonomous robot to treat. In addition to that, human society also has social and semiotic diversity based on a variety of regions and cultures. Many social rules and customs which are followed by people are neither universal nor constant. Meanings of words, gestures, and implications between speech lines vary amongst each families or communities even within the same regions. This being the case, robots existing in such a semiologically diverse world have no way of knowing meaning, value, or delivery for symbols beforehand, and as such are called upon to learn and acquire these things after being designed.

1.1 Acquisition of response strategy

In a continuous real world, social robots are required to obtain several behaviors including gestures autonomously. Taniguchi et al. proposed self-organizational imitation learning architecture which enables robots to acquire motion primitives by observing human unsegmented motion by combining switching auto-regressive model and keyword extraction method [T.Taniguchi et al. (2008)]. Tani et al. made humanoid robot to acuire several motions in self-organizing way by using RNNPB (Recurrent Neural Network with Parametric Bias) [Tani et al. (2004)]. In addition to that, robots living with humans in society should adopt a response strategy for expected actions corresponding to human behaviors in a real time manner. The changing of behavioral rules in an interactive manner with humans is important to build the kind of relationships between robots and users. Many people playing with pet robots could expect such a dynamic learning process during interactions with the robots.

Brazeal et al. developed a facial robot, Kismet, which can interact with a human participants by regulating its behaviors based on motivation system in a real time manner. However, it did not acquire a variety of response strategies [Breazeal and Scassellati (2000)]. Several reinforcement learning and evolutionary approaches of acquiring response strategy have been taken into consideration [Mitsunaga et al. (2005)]. However, these methods are to maximize evaluation functions by requiring trial and error to human users. These have a large possibility of overloading the users by requiring much time to train a robot. Whenever a user tries to make a robot learn something through interaction, it is expected that learning should be achieved at least on a scale of minutes, using concise, uninterrupted, real-time continuous interactions. Ogata et al. developed a robot which can learn in a coadaptive way through human-robot interaction achiveing navigation task [Tetsuya et al. (2005)]. An RNNPB embedded in the robot acquires motion primitives and collaborated with an interacting user. Kubota also developed similar collaborative robot with a modular NN [Kubota et al. (2003)]. In these researches turn taking in the interactions occurs implicitly. In continuous interactions, turn is considered as a role in the collaborative task.

⋆ This paper was partially supported by a research grant from the National Institute of Informatics, by the Japanese Ministry of Education, Culture, Sports, Science and Technology, under a Grant-in-Aid for Young Scientists (Start-up) Project No. 20800060 and Grant-in-Aid for Creative Scientific Research No. 19GS0208.

In this paper, we propose a framework in which a robot can acquire response strategy in a real-time manner, focusing on the role reversal imitation occasionally seen in the behavior of infant children. That enables an autonomous robot to acquire simple interaction rules through human-robot real-time interaction in continuous time.

1.2 Role reversal imitation

Tomasello uses the term, "Role Reversal Imitation" to explain a part of children's basic social capability of language (symbols) acquisition [Tomasello (2000)]. The role reversal imitation is to reverse and imitate the roles to achieve a collaborative act in interactive communication. In bilateral interaction, when a person says "Bye-bye" in response to somebody's "Bye-bye", the couple of actions could be taken as a collaborative act. By reversing the mutual roles in collaborative act, the learner can grasp how to use the signs in social communication. In addition to that, each side must comprehends the role of his/her action in building the collaborative act and change those roles to continue social communication dynamically and smoothly.

Hinoshita et al. developed a couple of robots using MTRNN(Multiple Timescale Recurrent Neural Network) which interact with voice and motion and showed interaction rule emerges. Kuriyama et al. proposed the scheme of acquiring response strategy adaptively by utilizing the transfer entropy, focusing on contingency found in the interaction between an infant child and its parents [Kuriyama and Kuniyoshi (2008)]. These acquisition schemes are however, based on heuristic model. To understand the computational process of role reversal imitation, constructing parametric model which describes role reversal imitation process in a real time manner is important. In this paper, we propose a novel scheme of acquiring the role reverse imitation utilizing the EM (Expectation Maximization) algorithm.

2. RESPONSE STRATEGY ACQUISITION BY USING THE EM ALGORITHM

2.1 definition of roles

Many interactions regarded as bilateral collaborative tasks are seen to consist of autonomous actions and counteractive responses to them. In this research, the former action is referred to as "Initiative Action" and the latter as "Reactive Action". In this research, "action" is a concept incorporating these two actions. Hereinafter they are abbreviated as IA and RA, respectively [1]. The conductor of IA is called the "Initiator", while the conductor of RA is the "Reactor". To realize collaborative action, it is necessary that the two participants output actions while appropriately sharing roles between these two. In this research, we call reversing the roles of Initiator and Reactor "role reversal", while each participant's scheme of outputting these actions is called "response strategy".

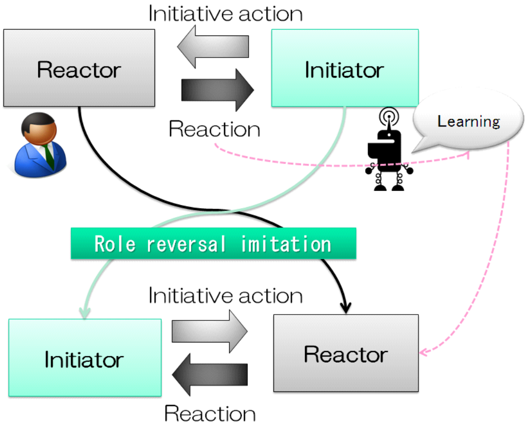

Fig. 1. Conceptual diagram of role reversal imitation

If RA and IA are explicitly linked, learning scheme to obtain response strategy through role reversal imitation is easily developed in the system. However, it becomes much more difficult in a system in which action is occured in continuous time and roles are not explicitly defined. In this paper, we develop a probability generative model in which RA respond to IA probabilistically. We construct the response strategy acquisition machine learning method for robots through real-time interactions with humans without separating the robot learning phase and behavior phase based on the generative model. The probabilittic model also has latent variables which represent roles [2]. By explicitly assuming participants' roles based on the generative model, the learning method realizes such a contiguous learning process. The conceptual diagram of role reversal imitation is shown in Figure 1.

2.2 Reaction Occurrence Probability

An overview of RA generative model is described in figure 2. The set of robot actions is taken as $A = a_1, a_2, \cdots, a_m$ and the set of human actions is $B = b_1, b_2, \cdots, b_m$. The sets added "null action" a_ϕ, b_ϕ are placed as A' and B'. Then a bijective function c exists to physically relate the human action and robot action and bilateral relations are generated through $c(a_i) = b_i$ [3]. In this research, we assume that RA with time delay occurs in stochastic manner caused by IA. In case of Time s, probability of RA occurring caused by $IA\ a_i$, $W^a_{a_i}, s$ is called a reaction occurrence probability, and it is defined by probability functions which decrease over time. Hereinafter, the variables are calculated according to updates for each step sequentially. In the formula below, s is a discrete-time step resulting from the sampling of continuous time for every interval time of $\triangle t$.

$$W^a_{a_i,s+1} \leftarrow \begin{cases} W^a_{a_i,s} + \dfrac{\Delta t}{\tau} & (IA\ a_i\ is\ observed) \\ \gamma W^a_{a_i,s} & (IA\ a_i\ is\ not\ observed) \\ \lambda & (i = \phi) \end{cases} \quad (1)$$

$$\gamma = \exp(-\frac{\Delta t}{\tau}) \quad (2)$$

[1] For voluntary actions, there are actions that request a counterpart's response and actions that do not. In this paper however, IA includes all non-RA actions responding to some particular actions.

[2] These indicate whether it is an Initiator or a Reactor role.

[3] In this regard, the correspondence problem of relation between an imitator and the imitatee is not covered in this research.

τ is a time constant proportional to the time interval from IA occurrence to RA occurrence. RA is generated from IA based on this probability, referred to as W^a. By regarding actions caused by null-action a_θ, b_θ as IA, the other IA can also be treated as a type of RA. IA occurs as a RA to a null-action in the probability of every step λ. There are two kinds of parameters, λ, τ, in these equations which are calculated by maximizing mutual information, shown later in this paper. IA only increases W^a. We need to indentify whether a_i of each time is an IA or not, in order to calculate W^a, but the variable, whether a_i is IA or not, is in general impossible to be observed.

2.3 Estimating reaction selection probability based on the EM algorithm

After RA occurrence is identified with reaction occurence probability, the specific RA, which is output based on IA, is selected probabilitisticaly. This probability $Z(b_j, a_i)$ of generating RA a_j which is a result from IA a_i is called "Reaction selection probability". A robot learns the action selection probability through the EM algorithm by observing human's RA, which is described in figure 1. We assume that the interaction history between robot and human for a certain time, $A^O = \{a(s) | a(s) \in A, s \in S_a\}$, and $B^O = \{b(s) | b(s) \in B, s \in S_b\}$ is taken. S_a and S_b thereby represent the set of times when the human and the robot respectively take action. If the robot learns the human response strategy, the robot should estimate human reaction selection probability Z. The log likelihood L becomes as follows.

$$L = \sum_{s \in S_b} \log(\sum_{a_i \in A'} W^a_{a_i,s} Z(b(s), a_i)) \tag{3}$$

In the formula above, $b(s)$ is an action selected by a human in time s. The robot needs to estimate the human reaction selection probability by maximizing this L and needs to share the human's reaction selection probability to obtain response strategy by reversing it as well. In this case, which B^O is brought about by which action of A^O can not be observed. Here, the problem maximizing the log likelihood cannnot be solved analitically. This is a problem. Such likelihood maximization problem can be solved by utilizing the EM algorithm. Function Q is defined as shown below.

$$Q(Z|Z^{(t)}) = \sum_{s \in S_b} \sum_{a_i \in A'} P_b(a_i | b(s), s) \log(Z(b(s), a_i))$$

$$P_b(a_i | b(s), s) = \frac{W^a_{a_i,s} Z^{(t)}(b(s), a_i)}{\sum_{a_j \in A'} W^a_{a_j,s} Z^{(t)}(b(s), a_j)} \tag{4}$$

Here, $Z^{(t)}$ is a parameter estimated after repeating t-times, while $P_b(a_i | b(s), s)$ is the probability that $b(s)$ is caused by a_i in the time s. This corresponds to E-step in EM algorithm. Variables are updated as follows:

$$M^{(t+1)}(b_j, a_i) = \sum_{s \in S_{b_j}} P_b(a_i | b_j, s) \tag{5}$$

$$Z^{(t+1)}(b_j, a_i) = \frac{M^{(t+1)}(b_j, a_i)}{\sum_j M^{(t+1)}(b_j, a_i)} \tag{6}$$

$$J^{t+1}(b_j, a_i) = \frac{M^{(t+1)}(b_j, a_i)}{\sum_{i,j} M^{(t+1)}(b_j, a_i)} \tag{7}$$

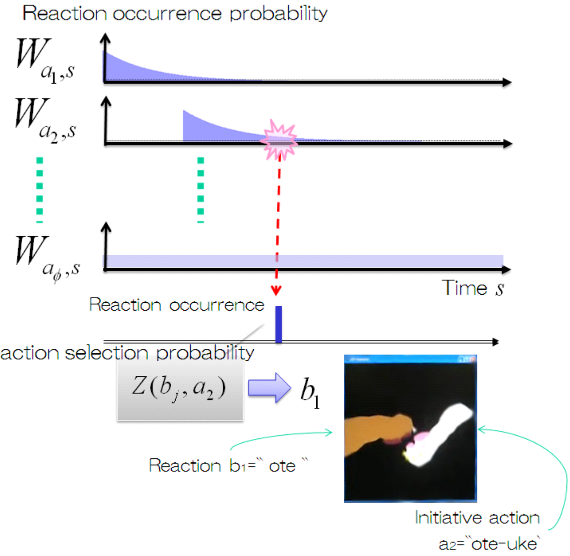

Fig. 2. Overview of probability generation model from Initiative action to Reactive action

In the formulas above, S_{b_j} is a subset of S_b which output b_j. $M(t+1)$ deserves counting frequency of IA-RA relations, while $J(t+1)$ is an estimate of joint probability distribution.

2.4 Role estimate and reaction output probability estimate

When W^a is determined, the reaction selection probability can be asymptotically estimated with the EM algorithm. However, it is not clear which action is determined as IA in an interaction with dynamically turn-takings, and as a result W^a cannot be deterministically decided. In this case the EM algorithm may help to simultaneously identify IA action during E-step. Assuming that the robots s step action, $a(s)$, is a reaction to the human action, b_i, then,

$$P_a(b_i | a(s), s) = \frac{W^b_{s,b_i} Z(c(a(s)), c^{-1}(b_i))}{\sum_{b_j \in B'} W^b_{s,b_j} Z(c(a(s)), c^{-1}(b_i))} \tag{8}$$

Here, $P_a(b_i | a(s), s)$ is a probability that the robot's action $a(s)$ is caused by b_i.

Using this model

$$W^a_{a_i,s+1} \leftarrow \begin{cases} W^a_{a_i,s} + P_a(b_\phi | a(s), s) \dfrac{\Delta t}{\tau} & (a_i \textit{ is observed}) \\ \gamma W^a_{a_i,s} & (a_i \textit{ is not observed}) \\ \lambda & (i = \phi) \end{cases}$$

$$W^b_{b_i,s+1} \leftarrow \begin{cases} W^b_{b_i,s} + P_b(a_\phi | b(s), s) \dfrac{\Delta t}{\tau} & (b_i \textit{ is observed}) \\ \gamma W^b_{b_i,s} & (b_i \textit{ is not observed}) \\ \lambda & (i = \phi) \end{cases}$$

Reaction output probability for each time is consecutively estimated to update the EM algorithm. It is also possible to estimate whether each action is recognized as IA or RA by observing $P_a(b_i | a(s), s)$, $P_b(a_i | b(s), s)$. In other words,

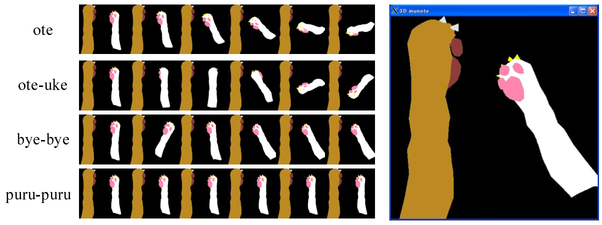

Fig. 3. Simulation screen and prepared unit of action

if $\phi = \text{argmax}_i P_b(a_i|b(s), s)$ is confirmed, $b(s)$ can be estimated as IA. [4]

2.5 Action selection by role reversal

Based on the scheme stated earlier, the robot determines how to select the action by estimating the human reaction selection rule, $Z(b, a)$ as $Z \leftarrow Z^{(t)}$. t is an iterative count. By using this estimated selection rule of human reactions and function c, reaction a_i to an initiative action b_j is selected in the following probability formula.

$$P(a_i|b_j) = \hat{Z}(c(a_i)|c^{-1}(b_j)) \tag{9}$$

The final probability of selecting a_i in the time step s is generated as below.

$$P(a_i|s) = \sum_{b_j \in B'} W^b_{b_j,s} \hat{Z}(c(a_i)|c^{-1}(b_j)) \tag{10}$$

3. EXPERIMENT

To put the effectiveness of the scheme that we propose in this research into context, we undertook the experiment using a robot which is equipped with computing interactions as follows.

3.1 Conditions of experiments

In this experiment, we set up the interaction system in the computer, as shown in figure 3, by preparing the paws of two dogs, which are handled by a user and a robot (learning program). The paw in front is handled by the user, while the paw on the far side is handled by the robot. The user outputs a pre-designated action by typing on the keyboard. Actions are allocated, as a_1 = "ote (give your paw)", a_2 = "ote-uke (hand to receive paw)", a_3="bye-bye", a_4 = "puru-puru" (swing). In this experiment, hands on both sides have the same action set, so we set b_i as a_i. Keys on a keyboard from 1 to 4 are respectively assigned as a_1 to a_4. When the user presses a key, the related action will begin. In this experiment, the robot can instantly recognize the action related to the key pressed without any false recognition. The frame rate is set at 10[Hz]. The participants were requested that they should make the robot learn

(1) action of "ote" (giving a paw) to the user's "ote-uke" (hand to receive)
(2) action of returning bye-bye to the user's bye-bye

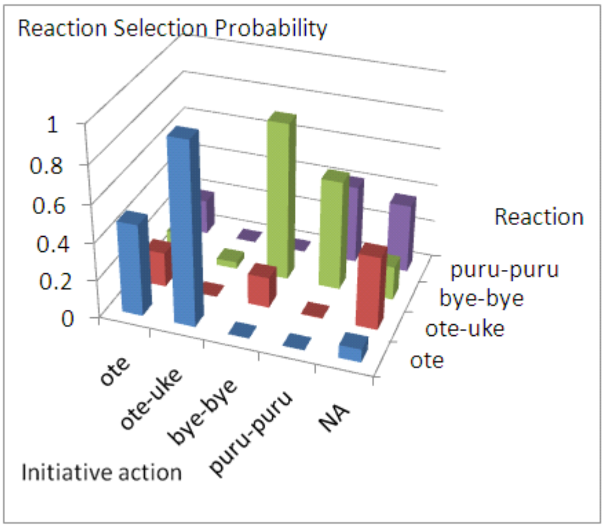

Fig. 4. Reaction selection probability $\hat{Z}$ after learning

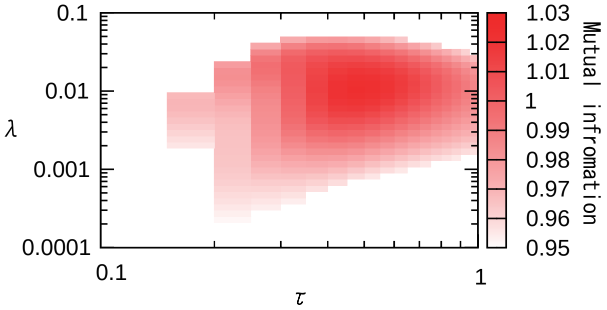

Fig. 5. Transition of mutual information by λ, τ

Therefore, the participatns gave "ote" as a reaction to the robot's "ote-uke', and "bye-bye" to the user's bye-bye, while taking some non-ruled actions in other cases. When the robot output actions, we did not take the generation model shown in formula 10 on how to output the robot action in this experiment for the purpose of targeting acquisition process and role estimate. In order to focus on parameter estimate of response strategy and role estimate, we instead took a generation scheme to select reaction based on the reaction selection probability by making robot output a reaction every 2[s] in average.

3.2 Experiment results

During the 300 seconds that the interactions are performed, the user does his best to respond to the action described in the experiment conditions. The EM algorithm was tested five times repeatedly with the conditions of $\tau = 0.2$ and $\lambda = 0.001$. The resulting reaction selection probability $Z(b_j, a_i)$ is shown in the figure 4. The result shows that the response strategy of reversing the "ote" and "bye-bye" actions from the user was properly acquired. Also, the mutual information I which was searched to maximize the value of λ, τ is calculated with the joint probability $J(b, a)$.

$$I(B, A) = \sum_{b \in B} \sum_{a \in A'} J(b, a) \log \frac{J(b, a)}{J(b)J(a)} \tag{11}$$

[4] The estimation $W_{a_j,s}$ here is taken in accordance with information up to step s. In this paper we adopt this estimate value, albeit no so strictly.

step	Subject	Reaction	Estimated corresponding initiative action
613	User	ote-uke	NA
618	Robot	ote	ote-uke
624	Robot	bye-bye	NA
630	User	bye-bye	bye-bye
1326	Robot	ote-uke	NA
1329	Robot	ote	NA
1331	User	ote	ote-uke

Table 1. Parts of role estimate result

As seen in the figure 5, the mutual information was maximized at around $\lambda = 0.01$, $\tau = 0.5$. This means that the robot understand the human-robot interaction relate in the informational sense.

In order to understand that, both interaction histories, A^O and B^O, are generated in continuous time with the highest joint probability. This information identifies the extent to which RA is delayed after IA occurrence. As it is difficult to decide the actual role in an interaction, the measurable judgment has not been made yet, but the typical sample of the experiment result is shown in the table 3-2. The table shows (1) as an estimate of the step 613- 630, and (2) the step 1326 - 1331. When no action has occurred, it is not displayed here. In the action steps (1), initially the user shows "ote-uke", and then the robot gives hand of "ote", providing the estimate that the user's "ote-uke" is "IA" and the robot "ote" is a reaction to the user's action. It also shows that 0.6 seconds after that, the robot performs the "bye-bye" action, which is not estimated as a reaction, but as IA. Next, the user's' bye-bye" is estimated as a reaction to the robot's "bye-bye". The flow of these actions helps us understand that turn-taking is being done in a real-time, and that role reversal is properly performed. In the action steps (2), the robot gives "ote" after "ote-uke", and 0.2 seconds after that, the user gives "ote". This means that the robot recognized that the "ote" by the users was the response to "ote-uke" the robot gave earlier, not to the robot's "ote". However, looking through the complete history of role estimates, we saw a tendency in which the human action taken just after the robot action is recognized as a reaction to the robot. This is because just after IA occurs, RA is generated most easily since W probability of reaction generation is defined as exponential distribution in a temporal axis. The actual RA to IA however, is not distributed as defined. It is considered to be distributed with a peak at a certain time delayed after IA occurrence. In this regard, we should examine changing the distribution model of W.

4. CONCLUSION

In this research, we proposed the scheme of acquiring the response strategy by the autonomous robot through the real-time interaction and role reverse imitation using limited motion primitives. This technique is structured scheme using the EM algorithm and is one in which the role of the action, the expressed action, and the corresponding IA to RA relationships are chiefly set as hidden variables. The effectiveness of this scheme is verified with interactions in virtual area structured on a computer system. We also proposed the method of estimating the dynamically changing roles in real-time interactions. Humans usually initiate changes in interactive relations, but by utilizing the role estimate method, it becomes possible for both humans and robots to produce more natural interactions. The next prospective step is to verify the effectiveness of this scheme by using an actual robot, and to leverage this scheme to achieve this not only in continuous time, but to expand it with continuous interaction.

ACKNOWLEDGEMENT

This research is supported by a Grant-in-Aid Creative Scientific Research 2007-2011 (19GS0208) funded by the Ministry of Education, Culture, Sports, Science and Technology, Japan. It is also partially supported by a research grant from the National Institute of Informatics.

REFERENCES

Asada, M., MacDorman, K., Ishiguro, H., and Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*, 37(2-3), 185–193.

Breazeal, C. and Scassellati, B. (2000). Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior*, 8(1), 49.

Kubota, N., Hisajima, D., Kojima, F., and Fukuda, T. (2003). Fuzzy and neural computing for communication of a partner robot. *Journal of muliple valued logic and soft computing*, 9(2), 221–239.

Kuriyama, T. and Kuniyoshi, Y. (2008). Acquisition of Human-Robot Interaction Rules via Imitation and Response Observation. In *Proceedings of the 10th international conference on Simulation of Adaptive Behavior: From Animals to Animats*, 467–476. Springer.

Mitsunaga, N., Smith, C., Kanda, T., Ishiguro, H., and Hagita, N. (2005). Robot behavior adaptation for human-robot interaction based on policy gradient reinforcement learning. 218–225.

Nehaniv, C. and Dautenhahn, K. (2002). The correspondence problem. In *Imitation in Animals and Artifacts*, 41–61. MIT Press.

Tani, J., Ito, M., and Sugita, Y. (2004). Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. *Neural Networks*, 17(8-9), 1273–1289.

Tetsuya, O., Shigeki, S., and Jun, T. (2005). Acquisition of motion primitives of robot in human-navigation task : Towards human-robot interaction based on. *Transactions of the Japanese Society for Artificial Intelligence*, 20, 188–196.

Tomasello, M. (2000). *The Cultural Origins of Human Cognition*. Harvard University Press.

T.Taniguchi, Iwahashi, N., Sugiura, K., and Sawaragi, T. (2008). Constructive approach to role-reversal imitation through unsegmented interactions. *Journal of Robotics and Mechatronics*, 20(4), 567–577.