

Acquiring Peekaboo Communication: Early Communication Model Based on Reward Prediction

Masaki Ogino, Tomomi Ooide, Ayako Watanabe and Minoru Asada

JST ERATO Asada Synergistic Intelligence Project, Graduate School of Engineering, Osaka University

2-1, Yamadaoka, Suita, Osaka 565-0871, Japan

[ogino,asada]@jeap.org, [tomomi.ooide,ayako.watanabe]@ams.eng.osaka-u.ac.jp

Abstract—Infants become sensitive to the regular behavior of their caregivers by the end of 4 months old. In this paper, we propose a communication system for a robot to acquire early communication. The acquisition of the communication is proceeded by the interactions of the three components; the memory module, the reward prediction module and the internal state module. The emotional change triggers the transfer of the sensor data stored in the short-term memory to the long-term memory. Once the memory segments are formed, the sensor data are compared with them. When the coincidence of the starting signal of stored data with the sensor data is detected, the prediction of the reward begins. The responses of the simulated robot with the proposed system are examined with and without the memory module when the caregiver takes the regular and irregular peekaboo communication. The results partly explain the behaviors observed in infants.

Index Terms—reward prediction, memory, emotion, peekaboo, early communication

I. INTRODUCTION

Recently, various kinds of robots such as humanoid robots and android have been developed that are intended to communicate with humans [10]. Such kinds of robots should be evidently endowed with an adaptive communication ability. On the other hand, the knowledge about human adaptability in communication is not sufficient for applying the theories to the communication 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 the humans develop adaptive communication ability [1].

From the day an infant is born, 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.

Infants studies show that infants become to sensitive to the regular behaviors of caregivers by 4 months after birth [6]. In the same month of development, infants begin to adjust the timing of the communication of their own mothers and, furthermore, generalize the timing so that they show the same response to unknown person whose timing is the same as their mother's [5]. When the interchange begins between infants and caregivers, the infants develop an ability to predict interactions with social partners [5].

Rochat et al. [6] investigate the responses of 2, 4 and 6 month infants to regular and irregular peekaboo communication. Two month old infants show equal attention and smile levels both to regular and irregular peekaboo communications. However, 4 and 6 month old infants show less attention and more smiles to regular peekaboo than to irregular peekaboo.

These experiments indicate that after 4 months (1) infants memorize the behavior of caregivers, and (2) adjust the timing with expectation of the next behavior based on the memory. We hypothesize that two emotional processes take important roles in these processes. The first process is the memorizing. In brain science, it is observed that the emotional stimulus of the amygdala effects to the pathway from the cortical to hippocampus, and the memory of the events before and after the stimulus is strengthened [4] [2]. The second process is the prediction of the reward. In brain science, it is well known that the dopamine neuron in basal ganglia takes an important role in reward prediction [11].

In this paper, we propose a model for a robot to acquire the new communication based on the reward prediction. The system keeps the sensor data in short term memory and puts into the long term memory when the value of internal state corresponding to the emotion is increased. Once the memory is formulated, the sensor data is compared with it and the robot expects the regular response of the caregiver. With this model, the robot can acquire the early communication, peekaboo.

In the following, first, the communication model of peekaboo is introduced with a proposed model. Then, the details of the each module in proposed system are explained. Next, the experimental results are shown, and finally discussion and the conclusions are shown.

II. PEEKABOO IN COMMUNICATION

A. communication model of peekaboo

The aim of this study is to propose a system to acquire an early communication. The system is supposed to be endowed with the learning abilities to learn the regular behavior and to adjust the timing of the communication. Peekaboo is a good example of such communication. In developmental psychology, peekaboo is treated as one of the

communications in which infants adjust to the emotions and affections of caregivers, and often used in experiments to examine the infant abilities to detect regular social behaviour in communications and to predict certain behaviors [6].

In this section, the acquisition and expectation system based on the reward prediction is proposed and it is explained how the system works in the example of peekaboo communication.

B. communication before memorization

Without the memory, the robot cannot compare the sensor data with anything. In this stage, the robot should detect what is a communication unit. However, it is a difficult problem to segment the certain communication unit from the sequential data. We propose a system in which the internal state (emotion) takes an important role in segmentation of situations.

Fig. 1 shows the process of the peekaboo communication when the robot does not have a memory of the communication. In this communication, the internal state of the robot changes as follows,

- 1) When the caregiver hides his/her face by the hand, the robot becomes surprised due to the sudden change.
- 2) When the caregiver says "peeka...", the robot becomes pleased to hear the caregiver's voice.
- 3) While the caregiver hides his/her face, the robot becomes bored with the no change of the situation.
- 4) When the face of the caregiver appears again with the voice "...boo", the internal state of the robot changes its values largely with the sudden appearance of the face with the loud voice.
- 5) The robot memorizes the sensor data for the certain period around the moment at which the internal state changes largely.

Fig. 2 shows the proposed system and how it works in this process. The system consists of three modules; memory module, internal state module and reward prediction module. In the communication process mentioned above, the sensor data are input to the memory module and the internal state modules (a). The sensor data for certain period as well as the internal state are memorized in the short-term-memory all the time like a flight records (b). When the internal state changes largely, the trigger to transfer the short-term-memory to long-term-memory is sent from the internal state module to the memory module (c), and the new memory is segmented and constituted (d).

C. communication after memorization

After the memory constitutions, the reaction of the robot changes as shown in Fig. 3.

- 1) When the caregiver starts to hide his/her face by the hand, the robot compares the sensor data with the

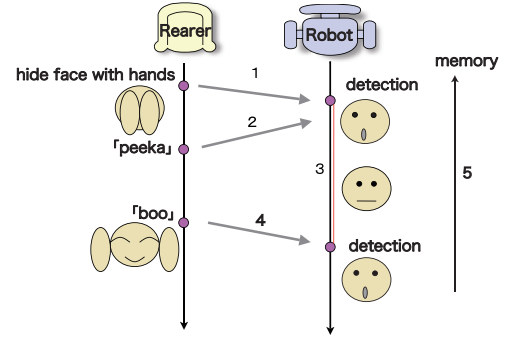


Fig. 1. The communication model in the case of no existing memory

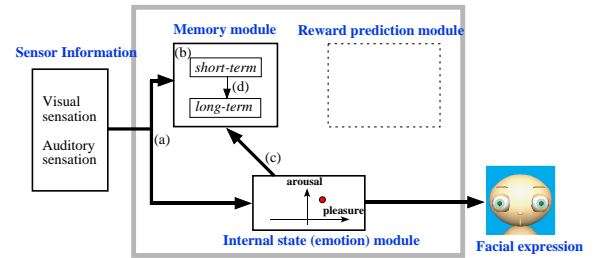


Fig. 2. The case of no existing memory

memorized data, and finds the memory that is similar to the current experience.

- 2) While the caregiver hides his/her face, the robot expects and predict the next behavior of the caregiver based on the memorized data.
- 3) When the face of the caregiver appears again with the voice "...boo", the robot compares the current sensor data with the memorized data, and changes its internal state based on the matching result.
- 4) The robot changes its facial expression based on its internal state.

In this process, the modules in the proposed system work as shown in Fig. 4. The sensor data are input to the memory module and internal state module as the previous process (a). The sensor data is all the time compared with the memorized data. When the matching is successful, the memory module sends the trigger to the reward prediction module which monitors the sensor data and predicts the future reward (b). Depending on the resultant difference between the sensor data and predicted one, the reward prediction module sends the error to the internal state module, which causes the increase (pleasure or surprise) or decrease (uncomfortableness) of internal state (c). Finally, the robot shows the facial expression based on the internal state (d). In the next section, the details of each module are explained.

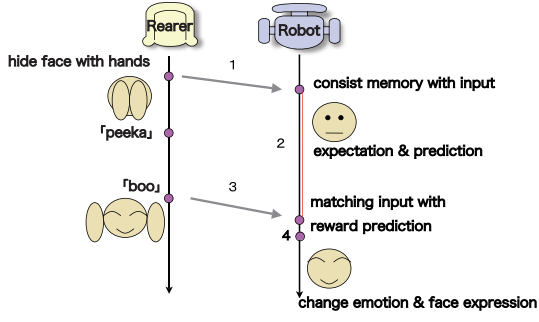


Fig. 3. The communication model of memory

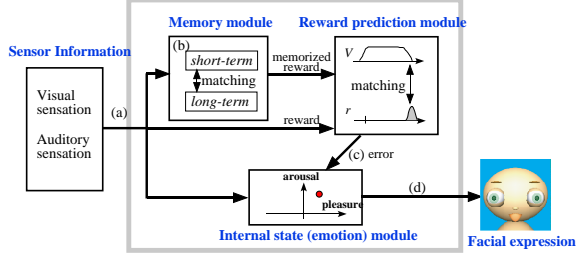


Fig. 4. The case of existing memory

III. ACQUIRING AND EXPECTING PEEKABOO

A. memory module

The memory module has two functions; memory formation and comparison. Both functions are realized by the short-term and long term memory. The short-term memory is the record of the sensor data, the internal state and the reward for certain time period. The short-term memory updates the record all the time: old data are overwritten by the new one.

When the internal state value exceeds the certain threshold, the internal state module triggers the formation of the memory: transition from the short-term memory to long-term memory. However, this transition is not just the copy of the short-term memory. When the recorded data in short memory contains another point with strong emotion (high value of the internal state), the long-term memory is formulated using the data from the previous point with strong emotion to the current point with strong emotion (Fig. 5).

The current data are compared with the long term memory segments. Here, only the first two sets of the data (the data at t and $t + 1$) are compared. When the matched memory are found, the following comparison starts and the data are compared all the time with the matched memory. The comparison is not exact matching. Some tolerance are permitted both in the value and the timing.

B. reward prediction

The properties of the dopamine neuron activity can be explained by the reward prediction. The neuron is activated when the reward comes. However, after the signal (the

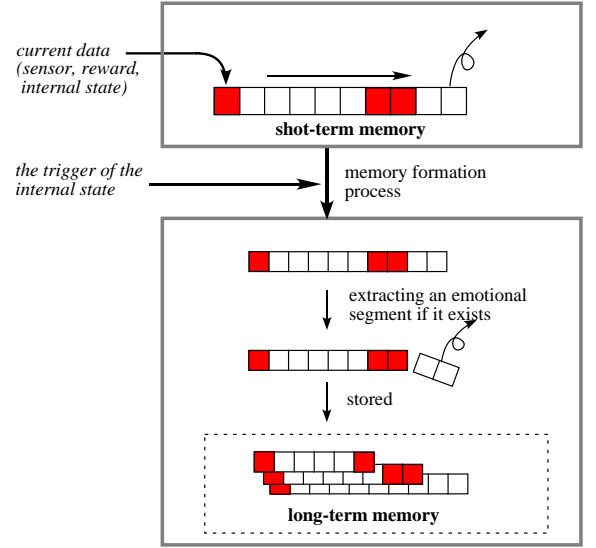


Fig. 5. Long-term memory formation

behavior or the event) that precedes the reward is learned, the dopamine neuron is activated when the signal is detected. These activation patterns are modeled well with the TD error in reinforcement learning [11] [8],

$$\delta(t) = r(t) + \gamma V(t) - V(t-1) \quad (1)$$

where $r(t)$, $V(t)$ and γ are the immediate reward, the estimated reward and the discount factor, respectively. Here, two kinds of reward $r(t)$ are calculated based on the sensor values,

$$r(t) = R_g(t) + R_s(t), \quad (2)$$

where R_g and R_s are the reward functions based on the image and the sound sensors, respectively. They are given as follows,

$$R_g(t) = \begin{cases} 1 & \text{when the caregiver's face appears} \\ 0 & \text{else} \end{cases} \quad (3)$$

$$R_s(t) = \begin{cases} 1 & \text{when the sound level exceeds the threshold} \\ 0 & \text{else} \end{cases}. \quad (4)$$

The estimated reward $V(t)$ is updated by the TD error learning,

$$V(t) \leftarrow V(t) + \alpha \delta(t). \quad (5)$$

Here, the discount factor γ is set to 1.0.

C. internal state

The internal state of a robot is constructed based on a human affect model. Numerous human affect models have been proposed. Among them, the circumplex model of the affect proposed by Russell [7] is adopted in our model because correspondence between the internal state

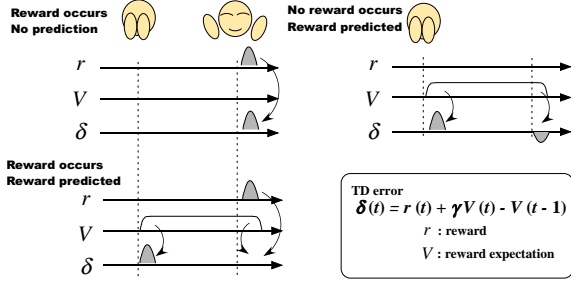


Fig. 6. Dopamine model

and facial expressions is relatively simple as explained later. In this model, it is assumed that the “arousal-sleep” axis and “pleasure-displeasure” axis are the most fundamental dimensions and that numerous different affect categories can be assigned to each as shown in Fig. 7.

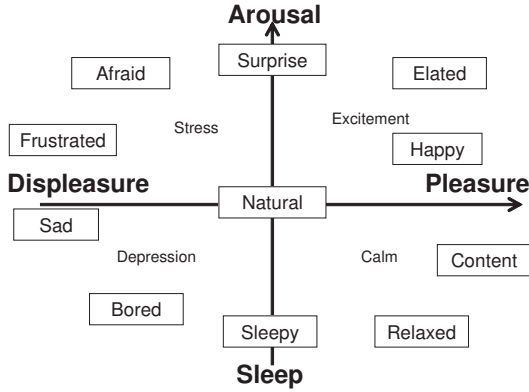


Fig. 7. Russell's two dimension emotion model

The robot's internal state S , consisting of two, independent variables, can be expressed as follows:

$$S = (p, a) \quad \begin{cases} (0 < p < 1) \\ (0 < a < 1) \end{cases}, \quad (6)$$

where p and a indicate the levels of pleasure and arousal, respectively, located in the space represented in Fig. 8. The corresponding facial expressions are shown in Fig. 8. The robot's internal state changes in the form of a relaxation dynamic equation:

$$\tau \dot{S} = -S + \Sigma r_j + S_0, \quad (7)$$

$$\Sigma r_j = r_i + r_e, \quad (8)$$

where S_0 , τ , and r_j are the internal states for the original point, the decay time constant, the effect of the external stimulus, respectively. r_i and r_e are the variables that effect to the internal state when none of the stored memory segments are matched nor matched, respectively. r_e is modified by the

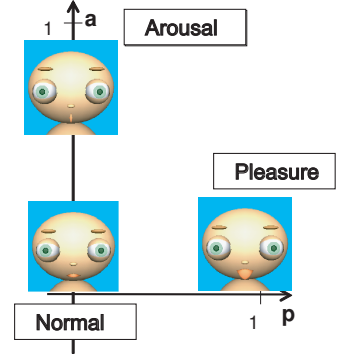


Fig. 8. Facial expression

error of the matched timing.

$$r_e = \Sigma e^{-\phi_x / \phi_0} r_x, \quad (9)$$

where x indicates the sensor type and ϕ is the timing error. The value of τ is changed so that the speed of the increase of the internal state becomes higher than that of decrease.

D. interest level

In addition to the internal state, the interest level is introduced that corresponds to the attention to the caregiver's face. This relates to the habituation that is one of the important factors to evaluate the infants behaviors in [6].

$$I = 1.0 - hab_m \times \frac{time}{at} \quad (10)$$

where $time$ is the time since the robot begins to attend to the caregiver's face. hab_m is the repeat counts of recalls of the same memory. at is the variable that changes depends on the detection of caregiver's face.

$$at = \begin{cases} 100 & (face \text{ appear}) \\ 150 & (face \text{ hide} \rightarrow face \text{ appear}) \\ 80 & (face \text{ appear} \rightarrow face \text{ hide}) \end{cases} \quad (11)$$

IV. EXPERIMENTAL RESULTS

To validate the proposed system, the simple communication experiments are examined.

The proposed system is implemented in a computer simulation. Fig. 9 shows the schema of the experimental environment. The robot is modeled as the computer graphics using OpenGL. The sensors that the robot has are the camera and microphone, and it shows the facial expression on the LC display based on its internal state as shown in Fig. 8. The facial images of the caregiver that are recorded in advance are used to evaluate whether the faces are hidden or not.

The examined communication experiment corresponds to the research of Rochat et al [6]. They examined the responses

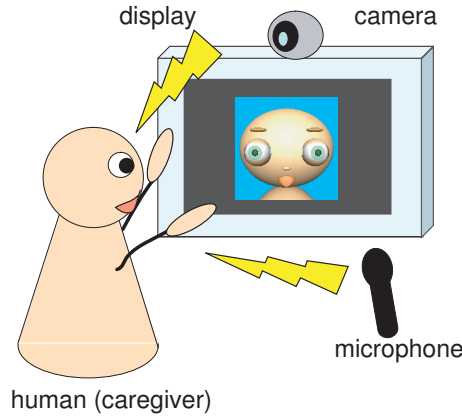


Fig. 9. Experimental environment

of infants aged 2, 4 and 6 month to regular peekaboo and irregular peekaboo. In the same way, we present the baby system the following peekaboo with and without memory modules,

1) *regular peekaboo*: The caregiver hides the face by the hands and waits some moments with saying "peeka...". After that, the caregiver exposes the face with saying "...boo". The caregiver repeats this behavior twice.

2) *irregular peekaboo*: The caregiver hides the face by the hands and exposes the face without saying "peekaboo". Or the caregiver only says "peekaboo" without hiding the face. The caregiver shows this abnormal peekaboo after the normal peekaboo.

Figs. 10-13 are the time courses of the interest level, the internal state, the reward, detection of the face, and detection of the sound. The images shown above the graphs are the input image to the robot and the facial expressions of the robot. The frame rate of the data is around 0.075 [sec]. The color near to blue indicates the low value, and the color near to pink indicates the high value. The values of the interest level, the internal state and the sound are continuous. On the other hand, the values of the reward and the detection of the face are binary. In detection of the face, "1" means that the face of the caregiver is detected, and 0 means not detected. The red and green lines indicate the starting and ending points of regular peekaboo, respectively. The yellow and blue lines indicate the starting and ending points of the irregular peekaboo, respectively.

Fig. 10 shows the responses of the baby system without the memory module to the regular peekaboo. The data show the responses to two peekaboo behaviors of the caregiver. Because the robot cannot form the memory and, therefore, cannot predict the caregiver's behavior, the responses of the emotions and the interest level show the same time courses.

Fig. 11 shows the responses of the baby system without the memory module to the irregular peekaboo. Firstly, the normal peekaboo is given, then two abnormal peekaboos

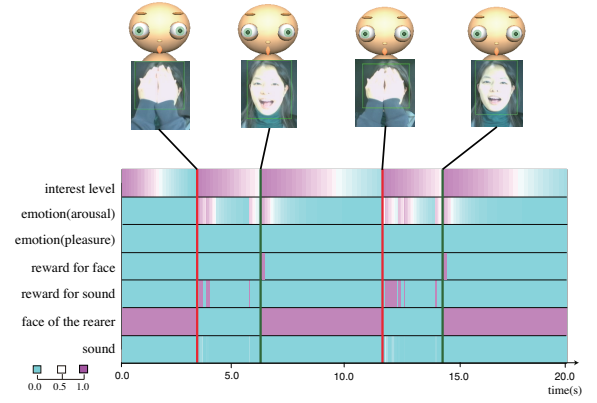


Fig. 10. The responses to the regular peekaboo without the memory module

(peekaboo without voice and peekaboo without hiding the face) are given. The responses do not change regardless of the normal and the abnormal peekaboos, because the robot cannot compare the communications. These results correspond to the observations of Rochat et al. [6], in which two months aged infants show the same level of attention and wondering both to the regular and irregular peekaboo.

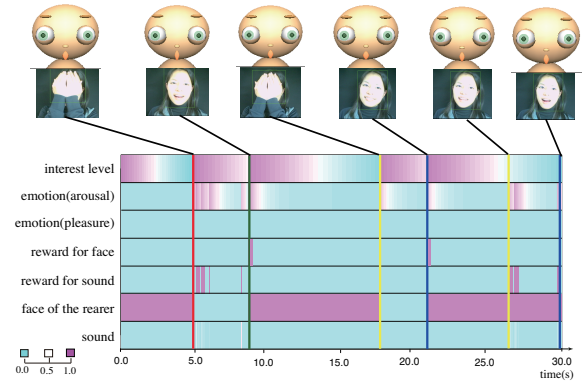


Fig. 11. The responses to the irregular peekaboo without the memory module

Fig. 12 shows the responses of the system with the memory module to the regular peekaboo. The robot shows the arousal to the first peekaboo, while it shows the pleasure to the second peekaboo thanks to the matching of the sensor data with the stored memory.

Fig. 13 shows the responses of the system with the memory module to the irregular peekaboo. In the second peekaboo, when the caregiver hides the face with the hands, the robot starts to predict the corresponding behavior of the caregiver. However, in this time, the caregiver does not show the "...boo" and keeps hiding the face. So, the robot shows the facial expression of surprise (red circle in the figure).

Rochat et al. show that the infants show more smiles and less attentions in regular peekaboo than in irregular

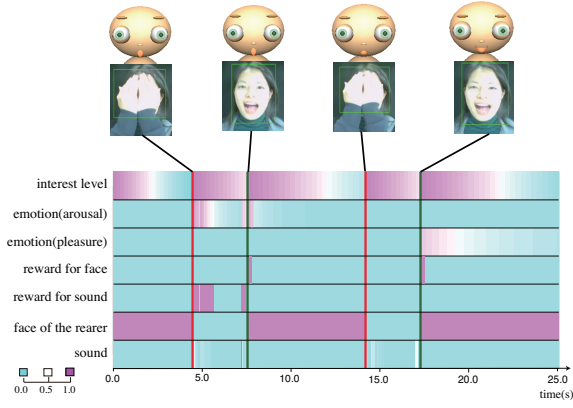


Fig. 12. The responses to the regular peekaboo with the memory module

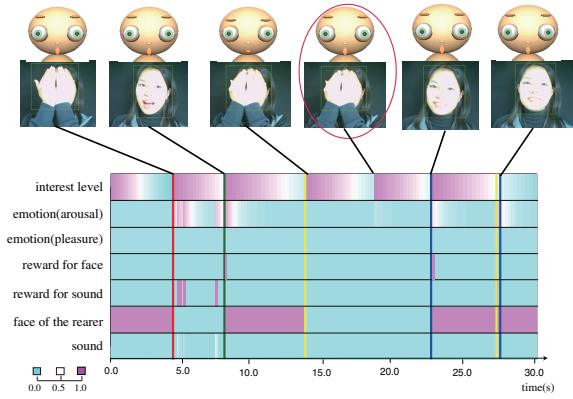


Fig. 13. The responses to the irregular peekaboo with the memory module

peekaboos. These results correspond to this observations.

V. DISCUSSIONS AND CONCLUSION

This paper proposes an early communication acquisition model that utilizes the reward prediction and the memory, both of which are closely related to the emotion (internal state). The simulated robot model with the proposed model show the similar responses to the regular and irregular peekaboo communications as the experiments in infants.

The proposed model is only the first stage to the early communication acquisition system. There are many important things left behind or too simplified. One of the most important issues is the memory model. In the current system, the sensor data are stored in the short term memory like the flight recorder, and the large emotion (internal state) change triggers the transfer of the short term memory to the long term memory. The experiment examined in the current system treat only one experience and did not test more general cases in which many memory segments could be generated and cause the problem of memory consolidation. The recurrent neural network model may have the superiority to consolidate the sequential data [9]. As the next stage, we would try the more

improved memory model so as to cope with the more general situations.

As Rochat et al. mentioned, infants show the different responses in peekaboo game in their developmental stages. In this paper, we model a peekaboo as the communication in which an infant passively observes the behavior of his/her caregiver. As the next stage, it is interesting to make a model that explains how infants begin to play a peekaboo game and to enjoy turn-taking with others. Mirza et al. already 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 [3]. This model might explain one of the contributing factors how infants begin to play the peekaboo game of their own motive. However, it is still an interesting question how the turn-taking game emerges based not only on the self experience but on the recognition of others. It is after 6th month that infants begin to play the peekaboo game, and it is also the same period that the shared attention and the imitation of behaviors begin.

The issue that we would like to approach next is how the emotion model of others affects to the acquisition of communication. In this paper, the pleasant level of the internal state increases when the caregiver's face appears or the reward is predicted well. However, the sympathy seems inevitably one of the important factors to evoke the communication. We are now planning to extend the proposed model to include the other's internal model and to examine how the communication acquisition changes with it.

REFERENCES

- [1] Cynthia Breazeal and Brian Scassellati. Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior*, 8(1):49–74, 2000.
- [2] James L. McGaugh. *MEMORY AND EMOTION*. Orion Publishing Group Ltd, 2003.
- [3] 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. *Adaptive Behaviour*, page to appear, 2007.
- [4] R. Paz, J.G. Pelletier, E.P. Bauer, and D. Pare. Emotion enhancement of memory via amygdala-driven facilitation of rhinal interactions. *Nature Neuroscience*, 9(10):1321–1329, 2006.
- [5] Philippe Rochat. *THE INFANT'S WORLD*, chapter 4. Harvard University Press, 2001.
- [6] Philippe Rochat, Jane G. Querido, and Tricia Striano. Emerging Sensitivity to the Timing and Structure of Protoconversation in Early Infancy. *Developmental Psychology*, 35(4):950–957, 1999.
- [7] J.A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 1980.
- [8] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.
- [9] Jun Tani and Jun Yamamoto. On the dynamics of robot exploration learning. *Cognitive Systems Research*, 3(3):459–470, 2002.
- [10] K. MacDorman, T. Minato, M. Shimada, S. Itakura, K. Lee, and H. Ishiguro. Evaluating humanlikeness by comparing responses elicited by an android and a person. In *Proceedings of the 2nd International Work-shop in Man-Machine Symbiotic Systems*, pages 373–383, 2004.
- [11] W. Schultz, P. Dayan, and P.F. Strick. A neural substrate of prediction and reward. *Science*, 275:236–250, 1997.