Human Friendship Estimation Model for Communication Robots

Shogo Nabe[1,2], Takayuki Kanda[1], Kazuo Hiraki[1,2], Hiroshi Ishiguro[1,3] and Norihiro Hagita[1,3]

[1]*ATR Intelligent Robotics Laboratories*
2-2-2 Hikaridai Seika-cho Soraku-gun, Kyoto, Japan
{nabe & kanda}@atr.jp
[2]*Department of General Systems Studies, Graduate School of Arts and Science, The University of Tokyo*
c/o Dr K Hiraki, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
khiraki@idea.c.u-tokyo.ac.jp
[3]*Osaka University*
Suita, Osaka, 565-0871, Japan
ishiguro@ams.eng.osaka-u.ac.jp

Abstract **– This paper reports about the analysis of inter-human interaction in the presence of a humanoid robot, which provides us the important non-verbal information for establishing relationship among humans and robots, as well as the model for communication robot to estimate friendships among humans around. Human-robot interactions recorded in videos are rich in useful information to develop the social abilities of communication robots, while there are many difficulties to analyze them since these videos are qualitative data. We applied the observation method, which is a common psychological method for analyzing qualitative data, to analyze the interaction among children and robot in an elementary school. We established a model to estimate friendships among children from their non-verbal interaction with each other; e.g. touching, distance, and gaze. Furthermore, we found a gender difference in their non-verbal interactions, and by separating the model for each gender, we achieved to discriminate friendly and non friendly relationship among children with 74.5% accuracy for male, and 83.8% for female.**

Index Terms – qualitative data, social robot, field trial, friendship estimation

I. Introduction

Recently there has been more and more research on communication robots taking part in our daily lives. These robots give people a strong feeling of presence with their physical body and human-like (or animal-like) way of interaction, and for this reason, they can play a physical, informative, and mental role [1] the way no other existing media can. Our final goal is to bring these robots into our daily lives, and for this purpose it is necessary to develop their fundamental ability to communicate with humans in a social way. Since little is known about what particular functions are needed to realize this human-robot communication, we believe it efficient to conduct an exploratory approach and try these robots in situations where people interact with the robots and with each other in our daily manner.

We have carried out a field trial with our communication robot "Robovie" at an elementary school for two months [2]. One of our goals was to elucidate how the interaction around robot would be carried out when it takes place in everyday situation. As a result, we found lively inter-human interaction in robot presence. For example, two girls being friends with each other, kept sticking together and often exchanged glances with smiles while playing with the robot. In another case of three girls, one of them kept acting towards robot as though she was keeping in step with the other two (**Fig. 1**). The girl touched the robot and spoke to it in the same way the other two did, but since the two were close friends with each other she did not seem to be interacting with them. However this situation came to an end with the robot starting a performance the three of them were very interested in, and during which they had an interaction that involved them all together.

We believe that it is important for communication robots to recognize the relationships among humans around them from two reasons. The first one is that we can spur human-robot interaction with this function. For instance, the robot with the two girls in the scene above may impress one girl, by saying "Aren't you with your friend today?" when she comes to play alone another time. The second reason is that we believe in the potential of communication robots to mediate the relationship between humans. The robot with the three girls in the scene above may recognize the relationships among them, and perform some behaviors to trigger interaction where all three can take part. To achieve this, it is necessary to utilize the non-verbal interaction between children.

The video data obtained in the field trials represents a rich deposit of information on non-verbal interaction among hu-

(a) Two among the three children are close friends　(b) Interaction among three triggered by the robot

Fig. 1: Inter-human interaction in robot presence

mans and robots, which we cannot get from controlled experiments in laboratories. We purposed to analyze it to retrieve the important non-verbal information for establishing relationship among humans and robots, as well as the model for communication robot to estimate friendships among human around. However, since this is qualitative data, there are difficulties to retrieve some findings from it. Therefore, it is a key issue to find a way to utilize this qualitative data.

In this paper, we report on our approach to establish a model of fundamental social ability for communication robots, by analyzing these data with the observation method. The observation method was established in the field of psychology. Through coding the data by hands, it enables us to make exploratory approach towards qualitative data and have the results in quantified manner. We refer to the fact, where Facial action coding system (FACS) led to a progress in image data processing. FACS developed by Ekman et al. [3], described facial expressions by focusing on muscle actions, and enabled us to recognize it through quantified data. Our approach towards the mount of video is the same; through quantifying the data with the observation method, we explore the possibilities to improve the robot's functions. We report a model to estimate friendships among children from occurring non-verbal interaction, such as touch, gaze, and distance. Even though the model for the time being needs the quantified data to be processed by hands, we believe that the finding from this model enables us to develop the software for social ability of communication robot.

Contrary to our approach for the social ability of communication robots, existing research works on this mainly focus on behavior between a single human and a robot, such as the joint-attention mechanism[4][5], and facial expressions[6]. However, little is done towards recognizing human relationship, in situations there are two or more people around the robot. On the other hand, there has been several research works in field of computer science that analyzed human relationship[7][8]. In real world, Choudhury visualized the amount of face-to-face conversation using wearable sensors [9]. While these works analyzed only symbolized information and did not mention what kind of relation the relationships they treated really are, our aim is to recognize true friendships among humans, through the natural non-verbal interaction friends have.

II. Field Trial

In our previous approach [2], we conducted a two month field trial in an elementary school, with the communication robot "Robovie". We observed the occurring human-robot interaction, and also tested our early model of friendship estimation. In our current study we will use the video data collected in this experiment to establish the friendship estimation model.

A. *Robovie and Its Person Identification*

Fig. 2 shows the communication robot "Robovie" [10] used in this experiment. The robot is capable of human-like expression and recognizes individuals by using various actuators and sensors. Its body was highly articulated to produce sufficient gestures for effective communication with humans. The sensory equipment consisting of auditory, tactile, ultrasonic, and vision sensors, and the processing and motor control hardware are located inside the robot's body.

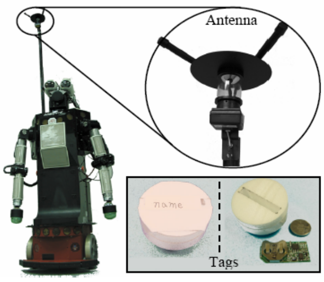

Fig. 2: Robovie and Wireless tags

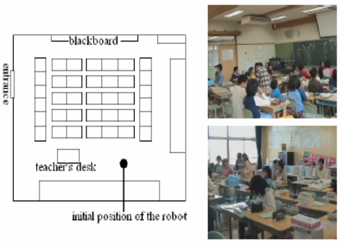

Fig. 3: Environment of the elementary school

This robot has a software mechanism for performing consistent interactive behaviors [11]. In design that the robot should communicate at a young child's level, one hundred interactive behaviors, such as shaking hands, hugging, exercising, kissing, singing, were developed. These interactive behaviors were shown based on some simple rules. These rules detect the sensory stimulus from children around, and decide the next interactive behavior.

To identify individuals, a wireless tag system capable of multi-person identification was installed. Recent radio frequency identification (RFID) technologies enable us to use contactless identification cards and chips on the field. In this study, children were given easy-to-wear nameplates (5 cm in diameter) in which a wireless tag was embedded. A tag (**Fig. 2**, lower-right) periodically transmitted its ID to the reader installed on the robot. This provided the robots with a robust means of identifying many children simultaneously.

B. *Experimental Setting and Results*

Using this communication robot we performed an experiment, which took place at an elementary school in Japan for two months. Subjects were 37 students (10-11 years old, 18 male and 19 female) who belonged to a fifth-grade class. The experiment lasted 2 months including 32 experiment days. We put the robot in a classroom (**Fig. 3**). The children were able to freely interact with the robot during a 30-minute recess after lunch time. In order to examine the estimation model, we handed out questionnaires to ask the children about their friendship within their classmates.

Although the detail of the results was reported in [2], here we briefly give an overview of the experiment, since the data will also be used in this research later. **Fig. 4** indicates the transition of interaction with children. The dotted lines separate the nine weeks during the two-month period. About ten children in total kept interacting with the robot every day (**Fig. 5-a**). The average interacting time of each child during the whole experiment was about 71 minutes; in the first two weeks the robot caused a big excitement, then the number of interacting children gradually decreased, and in the last two weeks it was increasing back again.

During the first 5 days, children made a crowd or form a line to play with the robot (**Fig. 5-b, c**). There was also a farewell party held for robot (**Fig. 5-d**) on the last day of experiment. While the previous study include these days in the data, our current one will exclude them, in order to deal with more common interaction around the robot.

III. FRIENDSHIP ESTIMATION MODEL

A. *Friendship Definition*

In this study, we consider the relationship between A and B as "friendship" if wither of them referred to the other as a friend on the handed questionnaire, which as the same definition as in our previous study [2][12]. However, we only deal with the relationships between same genders in this paper. It is known in psychology that children of this age prefer to hang out with children of same gender. It is assumed that there are little friendships between across genders comparing with those within a same, and even when existing, children might hesitate to reveal it in questionnaire. Indeed there were no such friendships reported in the questionnaires. What we aim to do is to establish a model to estimate human relationships, and we should not tackle these complicated problems in our early stage. We believe that estimating friendship within the genders will be the starting point to estimate cross-gender relationships in future works.

B. *Preparation for the model*

In contrast with our previous study, the friendship estimation model in this paper aims to exploit non-verbal interaction between friends. Our first approach towards this was to observe the interaction between children around the robot, and to reveal what interaction is available to estimate friendship. This pre-coding enabled us to code the categorized interaction and to quantify it. Finally we established a model to make estimation from the data. We used the data collected from our field trial in the elementary school, described in II, in the following sequence of analysis.

B-1. Pre-coding: reveal the interaction among children

[Method]

To reveal what kinds of interaction are available for friendship estimation, we conducted a preliminary analysis towards the video data of the experiment. We focused on individual children, and described each action they made. We selected a minute of scene, for each one of six randomly selected subjects (three males and three females), where they interacted with robot in presence of other children. The subjects were selected in the order they appeared on the video from the experiment starting from a middle day (17th day). This was to avoid the data from extremities, where the subjects have too much attention for the robot because of its novelty or because of its leaving the classroom. Analysis was done by describing all the "action" and "location towards the robot and other children" observed on the subjects.

[Results]

We obtained categorized interaction from this analysis as follows (**Table 1**, **Fig. 6**):

'Simultaneous stay': two children stay around the robot at the same moment (the friendship estimation model of our previous study [2] also used this information).

'Simultaneous appearance': two children come to play with the robot together.

'Simultaneous withdrawal': two children go away from the robot together.

'Together': two children positioning themselves very close to each other.

'Approaching': a child moves towards another child.

'Gaze': a child looks at the face of another child.

'Smile': a child smiles to another child.

'Touch': a child intentionally touches another child.

'Vocal interaction': either one speech, laughter, or other vocal action made in interaction between two children.

'Interaction in context': Considered as inter-human interaction because of its context, for example "a child touches the robot in the same place in the same way as another did just before".

B-2. Coding: qualify the non-verbal interaction

[Method]

Based on the found categories, we quantified the non-verbal interaction among children by coding the video. We coded 'simultaneous stay', 'simultaneous appearance', 'simultaneous withdrawal', 'touch', 'gaze', 'gaze with smile', 'together', 'approaching' among categories in **Table 1**, while we excluded 'vocal interaction' and 'interaction in context' from our analysis. 'Vocal interaction' was excluded since the video data were often too noisy that we could not figure out whether the children were uttering sounds or not. 'Interaction in context' was also excluded from the analysis because of the difficulty, to exploit it objectively.

The coding was done in interval-coding, as we sampled the children's presence and their positions compared to the robot's position every 10 seconds. During these 10 seconds, if a 'touch' or a 'gaze' was observed, it was also recorded (with its expression for 'gaze' too). Operational definition of each non-verbal interaction is shown in **Table 1**. Within these, 'gaze' counted the action between the two locating next to each other only. This was due to the fact that in the other cases, it was difficult to decide whether the gaze is towards the particular child, or towards something else (e.g. the robot between the two).

[Results]

Table 2 indicates the observed number and the occurrence frequency of each category of non-verbal interaction among friends and non friends. 'Occurrence frequency' is the rate of occurrence the particular category of interaction within ten seconds of 'simultaneous stay' between two, thus it is only defined for 'touch', 'gaze', 'smile', 'together', and 'approaching'. (Note that 'simultaneous appearance' and 'simultaneous withdrawal' are not the kind of interaction to occur *during* simultaneous stay.)

These results suggest that the categories enable us to estimate friendship from each of them. For 'simultaneous appearance' and 'simultaneous withdrawal', it was observed as often among friends as among non friends. There are,

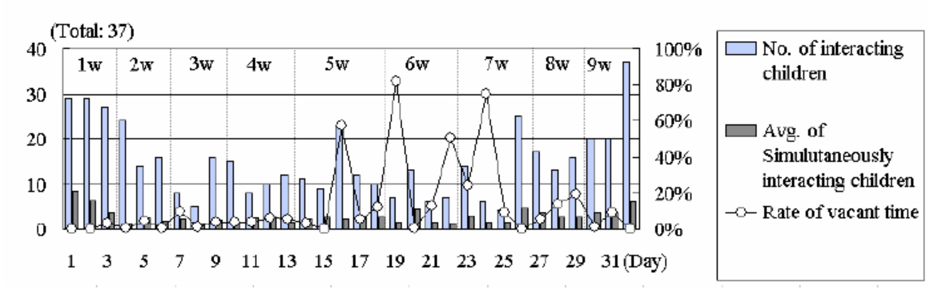

Fig. 4: Transitions of the interaction between children and the robot

(a) Daily interaction (b) Beginning of the first day

(c) Children formed a line (d) Farewell party

Fig. 5: Scenes of experiment

among friends as among non friends. There are, however, twice as many non friendly relationships as friendly ones, the results therefore suggest children preferred to appear or withdraw with friends rather than with non friends. For the categories where occurrence frequency could be defined, the results suggest a high performance of estimation if the values score relatively high for 'friends' comparing with 'non friends', which actually was the case.

In the data, pairs of females were nearly 4 times as often observed as the pairs of males. This is because females had longer interaction with the robot in general (this is also reported in [2]).

C. *Friendship Estimation Model*

From these results we designed a model to estimate friendships. All the non-verbal interactions coded above were used in the model.

We made the model to estimate the friendships out of children who interacted with the robot for more than 10 minutes. This led us to estimate 51 friendships (17 for males) out of 160 relationships (34 for males), where there are 106 friendships (52 for males) among 324 relationships (153 for males) between same-gender in the class. This is unavoidable since no model can estimate friendship between people who never appear within range of observation.

In the model, the estimated friendship between children A and B was defined as:

$$Friend(A,B) = if\left(Friendliness(A,B) > F_{TH}\right) \qquad (3)$$

$$Friendliness(A,B) = \sum_i \alpha_i * Interaction_i(A,B) \qquad (4)$$

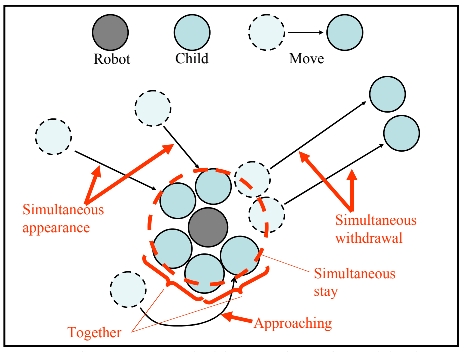

Fig. 6: non-verbal interaction of possitioning

$$\begin{aligned}
&= \alpha_{simul.stay} * Interaction_{simul.stay}(A,B) \\
&+ \alpha_{simul.appearance} * Interaction_{simul.appearance}(A,B) \\
&+ \alpha_{simul.withdrawal} * Interaction_{simul.withdrawal}(A,B) \\
&+ \alpha_{touch} * Interaction_{touch}(A,B) \\
&+ \alpha_{gaze} * Interaction_{gaze}(A,B) \\
&+ \alpha_{smile} * Interaction_{smile}(A,B) \\
&+ \alpha_{nearbying} * Interaction_{nearbying}(A,B) \\
&+ \alpha_{approaching} * Interaction_{approaching}(A,B)
\end{aligned}$$

where i corresponds to each non-verbal interaction. F_{th} is a threshold. $Interaction_i(A,B)$ is a function that gives a score based on the interaction components between A and B, which is normalized according to the expected occurrence rate for each child. α_i is a weight for each interaction to be considered in the model, which is the most essential part of this estimation.

IV. Findings from the Model

A. *Contribution of each Interaction in the Estimation*

[Method]

We conducted discrimination analysis on the coded data. This enabled us to obtain best α_i values (in (4)) to discriminate the relationships among this particular data, while it is not assured that these parameters are applicable for other data. Our aim is to establish a model that improves friendship estimation in a certain setting, and not to reveal in what particular setting

Table 1: Categories of interaction found in the pre-coding analysis of video data

Category	*operational definition*
Simultaneous stay time	Two children observed at the same moment
Simultaneous appearance	Two children appeared within 10 seconds
Simultaneous withdrawal	Two children went away within 10 seconds
Touch	A child touched another one by his/her palm (Do not include touch by other parts of body)
Gaze	Looking at a immediate neighbor's face
Smile	Smiling to an immediate neighbor.
Together	Two children separated by less than 50 cm.
Approaching	'Together' condition is a prequisite One child moved his/her position towards the other, while the other did not move his/her position.
verbal interaction	(not analyzed in coding)
interaction in context	(not analyzed in coding)

the model estimates the best. We consider that the obtained parameters reveal the type of interaction that is efficient for the friendship estimation, and that the performance of the model would reach its full potential, when the best way to piece together the information from the various non-verbal interactions is known. We settled on three parameter sets, which will perform best in estimating relationships between males, females, and both.

[Results]

Table 3 indicates the obtained values of α_i in (4). These results show which particular non-verbal interaction contributed to the estimation. Since each category of interactions has its data normalized, the values of α_i represent the meaningfulness of the corresponding information.

These results show the contributing categories in friendship estimation, where 'gaze', 'simultaneous withdrawal', 'touch' were the three most meaningful categories for males, while 'simultaneous stay' 'approaching' 'touch' were the ones for females. We consider that the categories with negative values contribute to the friendship estimation only when some interaction of other category has also occurred in high frequency in every relationship (e.g. girls often being close but seldom approaching to each other are likely to be friends). However it is not clear how these categories are linked to each other, and to reveal this is the challenge for our future work.

B. Evaluating the Model

[Method]

To evaluate the friendship estimation model we established, we applied two measures. A way to measure the model's performance that can come to mind is:

$$rate = \frac{number\ of\ relationships\ estimated\ correctly}{total\ number\ of\ relationships}$$

which we define as 'discrimination accuracy' in this paper. However, this measure does not work appropriately, when the number of friendships is relatively small against the whole relationships. For example, if we consider all the relationships among class, including male-female relationships, there would be 106 friendships among 666 relationships from the questionnaire response. If we suppose towards this data a classifier that classifies every relationship as non friends, the discrimination accuracy would rate 84.1%, which would mean the evaluation is completely useless.

In this regards, our previous study [2] proposed a second measure, which focuses on the relevance of the estimation ('coverage' and 'reliability'). This measure is defined as:

$$coverage = \frac{number\ of\ friendships\ estimated\ correctly}{number\ of\ friendships\ estimated}$$

$$reliability = \frac{number\ of\ friendships\ estimated\ correctly}{number\ of\ friendships\ from\ the\ questionnaire}$$

In this study, since the number of existing friendships is not so small compared to the whole number of relationships, it is possible to apply both 'discrimination accuracy' measure and 'coverage and reliability' measure. We apply the former method to evaluate the model as a whole, and the latter to see the trade-off between coverage and reliability in the model.

[Result]

Table 4 and **Fig. 7** indicate the result of the estimation using two models: the estimation model from our previous study, which uses 'simultaneous stay' time (the data were collected by video coding, not by RFID tags), and the estimation mode using the non-verbal interactions coded from the video. As mentioned in the method described in IV-A, each model has three sets of parameters for estimating relationships of both genders, male, and female. (Note that 'both gender' only counts relationships between same genders, as mentioned in III-A)

Table 4 indicates the discrimination accuracy of the established models. In this comparison the threshold of the model (in (3)) were set to a value ensuring the best performance from the parameter set. The model using all the information from non-verbal interaction had a discrimination accuracy of 71.3%, which improved the model using only 'simultaneous stay' by 6.3%. This improvement was more remarkable when the model estimates relationships within the same gender. In this case the model's discrimination accuracy increased by 23.6% for males and by 10.5% for females. This result suggests that using non-verbal interaction in friendship estimation is efficient, and that the performance improves by considering the difference in gender.

Fig. 7 shows the coverage-reliability chart. In this figure, *random* represents the reliability of random estimation, where we assume that all relationships are friendships (this will indicate the lower boundary of estimation). There is obviously a trade-off between reliability and coverage, which is controlled by F_{th}. Taking into account the non-verbal interactions improved the estimation of relationships within the same gender, given that the model successfully estimated 20% of the friendships with nearly 100% accuracy and 50% of them in between 60-80% accuracy.

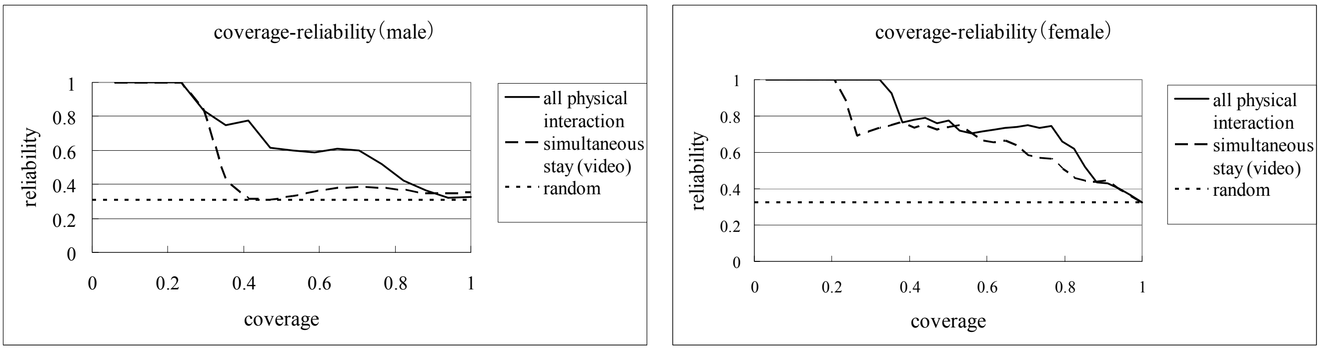

Fig. 7: Illustration of friendship estimation results

Table 2: Observed number and occurrence frequency of each non-verbal interaction

friends / non friends	Simul. stay		Simul. appearance		Simul. withdrawal		Touch		Gaze		Smile		Together		Approaching	
Num. of times	9660	11730	58	41	64	54	421	276	1431	1161	400	345	2058	1390	146	145
Occurrence freq.	---	---	---	---	---	---	0.044	0.024	0.148	0.099	0.042	0.03	0.214	0.12	0.015	0.012

Table 3: Weight of each non-verbal interaction (α_i) for friendship estimation models

	Simul. stay time	Simul. appearance	Simul. withdrawal	Touch	Gaze	Smile	Together	Approaching
Both gender	**1.13**	0.24	0.16	**0.40**	0.15	-0.13	0.28	**-0.30**
male	0.36	-0.12	**0.63**	**-0.44**	**1.33**	-0.23	0.15	-0.22
female	**2.10**	0.25	-0.29	**0.52**	-0.18	0.31	0.50	**-0.67**

Table 4: Performance of friendship estimation model

	Simul. time	All interaction
Both gender	65.0%	71.3%
Male	50.9%	74.5%
Female	73.3%	83.8%

V. Conclusion

In this paper, we presented our model for friendship estimation among children for communication robots. The model was established by analyzing video data of inter-human interaction using the observation method. We revealed of importance of particular non-verbal interaction between friends from the analysis. There existed a gender difference in this interaction, where 'gaze', 'simultaneous withdrawal', and 'touch' were essential in males' case whereas 'simultaneous stay', 'approaching', and 'touch' were in females'. From these findings, our model discriminated friendly and non friendly relationships among children with 74.5% accuracy for males, and 83.8% for females, which represents respectively 23.6% and 10.5% improvement compared with the existing model. Although the model uses hand-coded data, it will be useful for developing social ability of communication robots in the future. Moreover, we believe that these results demonstrate the importance of analyzing videos showing human interaction from the field trials for the development of communication robots.

Acknowledgment

This research was supported by the Ministry of Internal Affairs and Communications of Japan. The authors wish to thank Associate Professor Naoki Saiwaki at Nara Women's University for his valuable advice and helpful discussions.

References

[1] T. Shibata and K. Tanie, "Physical and affective interaction between human and mental commit robot," *Proc. IEEE Int. Conf. on Robotics and automation*, pp. 2572-2577, 2001

[2] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, "Friendly social robot that understands human's friendly relationships," *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2215-2222, 2004

[3] P. Ekman and W. V. Friesen, *Manual for the Facial Action Coding System and Action Unit Photographs,* Consulting Psychologists Press, 1978.

[4] B. Scassellati, "Investigating Models of Social Development Using a Humanoid Robot," *Biorobotics*, MIT Press, 2000

[5] H. Kozima and E. Vatikiotis-Bateson, "Communicative criteria for processing time/space-varying information," *Proc. IEEE Int. Workshop on Robot and Human Communication,* 2001

[6] C. Breazeal and B. Scassellati, "A context-dependent attention system for a social robot," *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI'99),* pp. 1146-1151, 1999

[7] J. D. Eveland and T. K. Bikson, "Evolving electronic communication networks: an empirical assessment," *Proc. of the 1986 ACM conference on Computer-supported cooperative work,* pp. 91-101, 1986

[8] S. Nomura, S. Oyama, T. Hayamizu, and T. Ishida, "Improvement of HITS Algorithm for Detecting Web Communities," *Int. Symposium on Applications and the Internet*, pp. 132-140, 2002

[9] T. Choudhury and A. Pentland, "Sensing Face-to-Face Communication Using the Sociometer," *IEEE Int. Conf. on Ubiquitous Computing*, 2003

[10] H. Ishiguro, T. Ono, M. Imai, and T. Kanda, "Development of an interactive humanoid robot "Robovie" -An interdisciplinary approach," *Robotics Research,*, Springer, pp. 179-191, 2003.

[11] T. Kanda, H. Ishiguro, M. Imai, T. Ono, and K. Mase, "A constructive approach for developing interactive humanoid robots," *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems,* pp. 1265-1270, 2002

[12] Takayuki Kanda and Hiroshi Ishiguro, "Reading human relationships from their interaction with an interactive humanoid robot," *Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE),* 2004