Body Movement Analysis of Human-Robot Interaction

Takayuki Kanda, Hiroshi Ishiguro, Michita Imai, and Tetsuo Ono
ATR Intelligent Robotics & Communication Laboratories
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan
kanda@atr.co.jp

Abstract

This paper presents a method for analyzing human-robot interaction by body movements. Future intelligent robots will communicate with humans and perform physical and communicative tasks to participate in daily life. A human-like body will provide an abundance of non-verbal information and enable us to smoothly communicate with the robot. To achieve this, we have developed a humanoid robot that autonomously interacts with humans by speaking and making gestures. It is used as a testbed for studying embodied communication. Our strategy is to analyze human-robot interaction in terms of body movements using a motion capturing system, which allows us to measure the body movements in detail. We have performed experiments to compare the body movements with subjective impressions of the robot. The results reveal the importance of well-coordinated behaviors and suggest a new analytical approach to human-robot interaction.

1 Introduction

Over the past several years, many humanoid robots such as Honda's [Hirai *et al.*, 1998] have been developed. We believe that in the not-too-distant future humanoid robots will interact with humans in our daily life. Their human-like bodies enable humans to intuitively understand their gestures and cause people to unconsciously behave as if they were communicating with humans [Kanda *et al.*, 2002a]. That is, if a humanoid robot effectively uses its body, people will naturally communicate with it. This could allow robots to perform communicative tasks in human society such as route guides.

Several researchers have investigated about social relationships between humans and robots. For example, Kismet was developed for studying early caregiver-infant interaction [Breazeal, 2001]. Also, a robot that stands in a line [Nakauchi *et al.*, 2002] and a robot that talks with multiple persons [Nakadai *et al.,* 2001] have been developed. Furthermore, various communicative behaviors using a robot's body have been discovered, such as a joint-attention mechanism [Scassellati *et al.*, 2000].

On the other hand, methods of analyzing social robots, especially with respect to human-robot interaction, are still lacking. To effectively develop any systems in general, it is essential to measure the systems' performance. For example, algorithms are compared with respect to time and memory, and mechanical systems are evaluated by speed and accuracy. Without analyzing current performance, we cannot argue advantages and problems. For social robots, no analysis method has yet been established, thus, it is vital to determine what types of measurements we can apply. Although questionnaire-based methods have been often used, they are rather subjective, static and obtrusive (that is, we would interrupt the interaction when we apply a questionnaire). Less commonly, human behaviors are employed for this purpose, such as distance [Hall, 1966], attitude [Reeves and Nass, 1996], eye gaze (often used in psychology), and synchronized behaviors [Ono *et al.*, 2001]. Although those methods are more difficult to apply, the results are more objective and dynamic. Sometimes, interactive systems observe human behaviors for synthesizing behaviors [Jebara and Pentland, 1999]. However, they are still fragments rather than a systematic analysis method applicable for human-robot interaction.

In this paper, we present our exploratory analysis of human-robot interaction. Our approach is to measure the body movement interaction between a humanoid robot and humans, and compare the results with traditional subjective evaluation. We have developed an interactive humanoid robot that has a human-like body as the testbed of this embodied communication. Furthermore, many interactive behaviors have been implemented. It encourages people to treat the robot as a human child. We employ a motion capturing system for measuring time and space accurately.

2 An Interactive Humanoid Robot

2.1 Hardware

Figures 1 and 2 display an interactive humanoid robot "Robovie," which is characterized by its human-like body expression and various sensors. The human-like body

---

[*] This research was supported in part by the Telecommunications Advancement Organization of Japan.

consists of eyes, a head and arms, which generate the complex body movements required for communication. The various sensors, such as auditory, tactile, ultrasonic, and visual, enable it to behave autonomously and to interact with humans. Furthermore, the robot satisfies mechanical requirements of autonomy. It includes all computational resources needed for processing the sensory data and for generating behaviors. It can continually operate for four hours with its battery power supply.

2.2 Software

Using its body and sensors, the robot performs diverse interactive behaviors with humans. Each behavior is generated by a *situated module,* each of which consists of *communicative units*. This implementation is based on a constructive approach [Kanda *et al.*, 2002b]: "combining as many simple behavior modules (*situated modules*) as possible." We believe that the complexity of the relations among appropriate behaviors enriches the interaction and creates perceived intelligence of the robot.

Communicative Unit

Previous works in cognitive science and psychology have highlighted the importance of eye contact and arm movement in communication. *Communicative units* are designed based on such knowledge to effectively use the robot's body, and each unit is a sensory-motor unit that realizes certain communicative behavior. For example, we have implemented "eye contact," "nod," "positional relationship," "joint attention (gaze and point at object)." When developers create a *situated module*, they combine the *communicative units* at first. Then, they supplement it with other sensory-motor units such as utterances and positional movements for particular interactive behaviors.

Situated Modules

In linguistics, an adjacency pair is a well-known term for a unit of conversation where the first expression of a pair requires the second expression to be of a certain type. For example, "greeting and response" and "question and answer" are considered pairs. We assume that embodied communication is materialized with a similar principle: the action-reaction pair. This involves certain pairs of actions and reactions that also include non-verbal expressions. The continuation of the pairs forms the communication between humans and a robot.

Although the action and reaction happen equally, the recognition ability provided by current computer science is not as powerful as that of humans. Thus, the robot takes the initiative and acts rather than reacting to humans actions. This allows the flow of communication to be maintained. Each *situated module* is designed to realize a certain action-reaction pair in a particular situation (Fig. 3), where a robot mainly takes an action and recognizes the humans' reaction. Since it produces a particular situation by itself, it can recognize humans' complex reactions under limited conditions; that is, it expects the human's reaction. This policy enables developers to easily implement many *situated modules*. On the other hand, when a human takes an action toward the robot, it recognizes the human's action and reacts by using reactive transition and *reactive modules*; that is, some of the *situated modules* can catch the human's initiating behaviors and interrupt its operations to react to them (as shown in Fig. 4: the second module TURN).

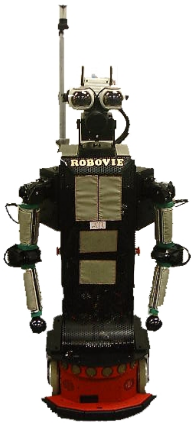

Figure 1: "Robovie"

Figure 2: Interactive behaviors

A *situated module* consists of *precondition*, *indication*, and *recognition parts* (Fig. 3). By executing the *precondition*, the robot checks whether the *situated module* is in an executable situation. For example, the *situated module* that performs a handshake is *executable* when a human is in front of the robot. By executing the *indication part*, the robot interacts with humans. With the handshake module, the robot says "Let's shake hands," and offers its hand. This behavior is implemented by combining *communicative units* of eye contact and positional relationships (it orients its body toward the human), and by supplementing a particular utterance ("Let's shake hands") and a particular body movement (offering its hand). The *recognition part* is designed to recognize several expected human reactions toward the robot's action. As for the handshake module, it can detect human handshake behavior if a human touches its offered hand.

The robot system sequentially executes *situated modules* (Fig. 4). At the end of the current *situated module* execution, it records the recognition result obtained by the recognition part, and progresses to the next executable *situated module*. The next module is determined by the results and the execution history of previous *situated modules*, which is similar to a state transition model.

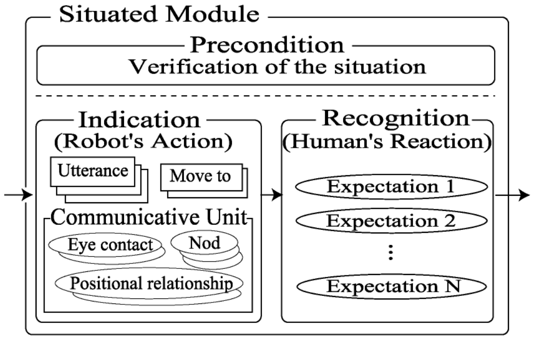

Figure 3: Situated module

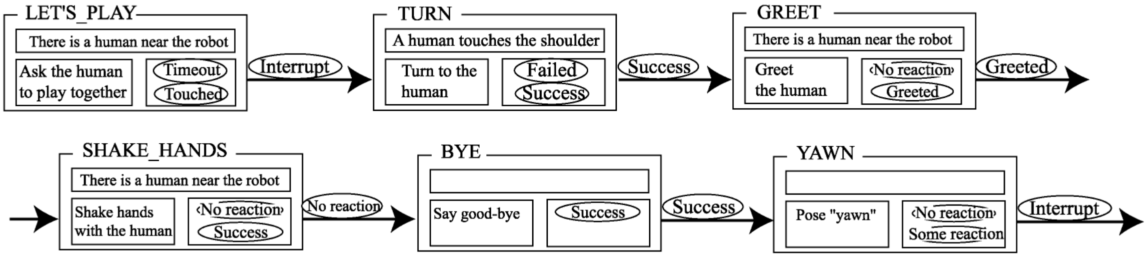

Figure 4: Example of situated module transition

2.3 Realized Interactive Behaviors

We installed this mechanism on "Robovie." The robot's task is to perform daily communication as children do. The number of developed *situated modules* has reached a hundred: 70 of which are interactive behaviors such as handshake (Fig. 2, upper-left), hugging (Fig. 2, upper-right), playing paper-scissors-rock (Fig. 2, lower-left), exercising (Fig. 2, lower-right), greeting, kissing, singing a song, short conversation, and pointing to an object in the surroundings; 20 are idling behaviors such as scratching its head, and folding its arms; and 10 are moving-around behaviors, such as pretending to patrol an area and going to watch an object in the surroundings.

Basically, the transition among the *situated* modules is implemented as follows: it sometimes asks humans for interaction by saying "Let's play, touch me," and exhibits idling and moving-around behaviors until a human acts in response; once a human reacts to the robot (touches or speaks), it starts and continues the friendly behaviors while the human reacts to these; when the human stops reacting, it stops the friendly behaviors, says "good bye" and re-starts its idling or moving-around behaviors.

3 Body Movement Analysis

3.1 Experiment Settings

We performed an experiment to investigate the interaction of body movements between the developed robot and a human. We used 26 university students (19 men and 7 women) as our subjects. Their average age was 19.9. First, they were shown examples how to use the robot, then they freely observed the robot for ten minutes in a rectangular room 7.5 m by 10 m. As described in section 2.3, the robot autonomously tries to interact with subjects. At the beginning of the free observation, the robot asks subjects to talk and play together, and then subjects usually start touching and talking.

After the experiment, subjects answered a questionnaire about their subjective evaluations of the robot with five adjective pairs shown in Table 1, which was compared with the body movements. We chose these adjective pairs because they had high loadings as evaluation factors for an interactive robot in a previous study [Kanda *et al.*, 2002a].

3.2 Measurement of Body Movements

We employed an optical motion capturing system to measure the body movements. The motion capturing system consisted of 12 pairs of infrared cameras and infrared lights and markers that reflect infrared signals. These cameras were set around the room. The system calculates each marker's 3-D position from all camera images. The system has high resolution in both time (120 Hz) and space (accuracy is 1 mm in the room)

As shown in Fig. 5, we attached ten markers to the heads (subjects wore a cap attached with markers), shoulders, necks, elbows, and wrists of both the robot and the subjects. By attaching markers to corresponding places on the robot and subjects, we could analyze the interaction of body movements. The three markers on the subjects' head detect the individual height, facing direction, and potential eye contact with the robot. The markers on the shoulders and neck are used to calculate the distance between the robot and subjects, and distance moved by them. The markers on the arms provide hand movement information (the relative positions of hands from the body) and the duration of synchronized movements (the period where the movements of hands of the subject and robot highly correlate). We also analyzed touching behaviors via an internal log of the robot's touch sensors.

3.3 Results

Comparison between the body movements and the subjective evaluations indicates meaningful correlation. From the experimental results, well-coordinated behaviors such as eye contact and synchronized arm movements proved to be important. This suggests that humans make evaluations based on their body movements.

Subjective Evaluation: "Evaluation Score"

The semantic differential method is applied to obtain subjective evaluations with a 1-to-7 scale, where 7 denotes the most positive point on the scale. Since we chose the adjective pairs that had high loadings as evaluation factors for an interactive robot, the results of all adjective pairs represent subjective evaluation of the robot. Thus, we calculated the *evaluation score* as the average of all adjective-pairs' scores. Table 1 indicates the adjective pairs used, the averages, and standard deviations.

Correlation between Body Movements and Subjective Impressions

Table 2 displays the measured body movements. Regarding eye contact, the average time was 328 seconds, which is more than half of the experiment time. Since the robot's eye height was 1.13 m and the average of subject

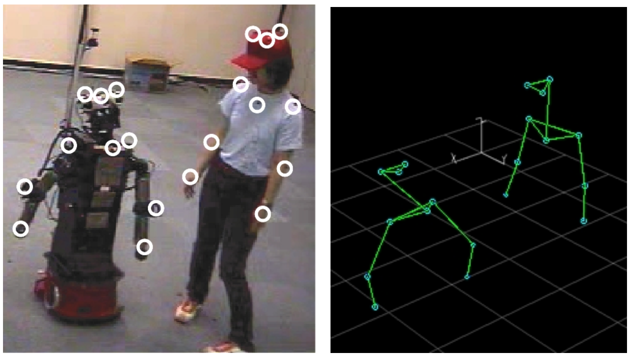

Figure 5: Attached markers (left) and obtained 3-D numerical position data of body movement (right)
In the left figure, white circles indicate the attached markers, and the circles in the right figure indicate the observed position of the markers

eye height was 1.55 m, which was less than their average standing eye height of 1.64 m, several subjects sat down or stooped to bring their eyes to the same height as the robot's. The distance moved was farther than what we expected, and it seemed that subjects were always moving little-by-little. For example, when the robot turned, the subjects would then correspondingly turn around the robot. Some subjects performed arm movements synchronized with the robot's behaviors, such as exercising.

Next, we calculated the correlation between the evaluation score and the body movements (Table 3). Since the number of subjects is 26, each correlation value whose absolute value is larger than 0.3297 is significant. We highlight these significant values with bold face in the table. From the calculated results, we found that eye contact and synchronized movements indicated higher significant correlations with the evaluation score.

According to the correlations among body movements, the following items showed significant correlations: eye contact – distance, eye contact – distance moved, synchronized behaviors – distance moved by hands, and synchronized behaviors – touch. However, these items (distance, distance moved, distance moved by hands, and touch) do not significantly correlate with the evaluation score. That is, only the well-coordinated behaviors correlate with the subjective evaluation. Isolated active body movements of subjects, such as standing near the robot, moving their hands energetically, and touching the robot repetitively, do not correlate to the subjective evaluation.

Estimation of Momentary Evaluation: "Entrainment Score"

The results indicate that there are correlations between subjective evaluation and body movements. We performed multiple linear regression analysis to estimate the evaluation score from the body movements, which confirms the above analysis and reveals how much each body movement affects the evaluation. We then applied the relations among body movements to estimate a momentary evaluation score called the *entrainment score.*

As a result of the multiple linear regression analysis, standardized partial regression coefficients were obtained, as shown in Table 4. The obtained multiple linear regression is as follows:

Adjective-pairs		Mean	Std. Dev.
Good	Bad	4.88	0.95
Kind	Cruel	4.85	1.29
Pretty	Ugly	5.08	0.93
Exciting	Dull	4.46	1.61
Likable	Unlikable	4.77	1.03
Evaluation score		4.81	0.92

Table 1: The adjective-pairs used for subjective evaluation, and the mean and resulting mean and standard deviation

	Mean	Std. Dev.
Distance (m)	0.547	0.103
Eye contact (s)	328	61.8
Eye height (m)	1.55	0.124
Distance moved (m)	35.2	17.0
Distance moved by hands (m)	108	29.5
Synchronized movements (s)	7.95	6.58
Touch (num. of times)	54.9	20.8

Table 2: Results for body movement

$$E = \alpha_{dist} \bullet DIST + \alpha_{ec} \bullet EC + \alpha_{eh} \bullet EH + \alpha_{dm} \bullet DM + \alpha_{dmh} \bullet DMH + \alpha_{sm} \bullet SM + \alpha_{touch} \bullet TOUCH + \alpha_{const} \quad (1)$$

where *DIST, EC, EH, DM, DMH, SM,* and *TOUCH* are the standardized values of the experimental results for the body movements. Since the evaluation was scored on a 1-to-7 scale, evaluation score E is between 1 and 7. The multiple correlation coefficient is 0.77, thus 59% of the evaluation score is explained by the regression. The validity of the regression is proved by analysis of variance ($F(7,18) = 3.71,\ P < 0.05$).

The coefficients (Table 4) also indicate the importance of well-coordinated behaviors. Eye contact and synchronized movements positively affected the evaluation score; on the contrary, distance, distance moved and touch seem to have negatively affected the evaluation score. In other words, the subjects who just actively did something (standing near the robot, moved around, and touched repeatedly), especially without cooperative behaviors, did not evaluate the robot highly.

Because we can momentarily observe all terms involved in the body movements of the regression (1), we can estimate a momentary evaluation score by using the same relations among body movements as follows:

$$E(t) = \alpha_{dist} \bullet DIST(t) + \alpha_{ec} \bullet EC(t) + \alpha_{eh} \bullet EH(t) + \alpha_{dm} \bullet DM(t) + \alpha_{dmh} \bullet DMH(t) + \alpha_{sm} \bullet SM(t) + \alpha_{touch} \bullet TOUCH(t) + \alpha_{const} \quad (2)$$

where designations such as *DIST(t)* are the momentary values of the body movements at time *t.* We named this momentary evaluation score the "*entrainment score*," with the idea that the robot entrains humans into interaction through its body movements and humans move their body according to their current evaluation of the robot. The

	Evaluation	Dist.	E. C.	E. H.	D.M.	D.M. H.	S. M.	Touch
Dist.	-0.04	1.00						
E.C.	**0.57**	**-0.47**	1.00					
E.H.	0.08	**-0.39**	0.29	1.00				
D.M.	-0.32	0.20	**-0.43**	0.02	1.00			
D.M.H.	0.01	-0.04	-0.21	-0.09	**0.49**	1.00		
S. M.	**0.54**	-0.05	0.28	-0.05	0.15	**0.61**	1.00	
Touch	0.21	**-0.45**	**0.49**	-0.07	-0.15	**0.35**	**0.41**	1.00

Table 3: Correlation between subjective evaluation and body movements. (E.C.: eye contact, E.H.: eye height, D. M.: distance moved, D.M.H.: distance moved by hands, S.M.: synchronized movements)

evaluation score and entrainment score satisfy the following equation, which represents our hypothesis that the evaluation forms during the interaction occurring through the exchange of body movements:

$$E = \int_0^t E(t) \Big/ t \qquad (3)$$

Let us show the validity of the estimation by examining the obtained entrainment score. Figure 6 shows the entrainment scores of two subjects. The horizontal axis indicates the time from start to end (600 seconds) of the experiments. The solid line indicates the entrainment score *E(t)*, while the colored region indicates the average of the entrainment score *E(t)* from the start to time *t* (this integration value grows the estimation of *E* at the end time).

The upper graph shows the score of the subject who interacted with the robot very well. She reported after the experiment that, “It seems that the robot really looked at me because of its eye motion. I nearly regard the robot as a human child that has an innocent personality.” This entrainment-score graph hovers around 5 and sometimes goes higher. This is because she talked to the robot while maintaining eye contact. She performed synchronized movements corresponding to the robot’s exercising behaviors, which caused the high value around 200 sec.

At the other extreme, the lower graph is for the subject who became embarrassed and had difficulty in interacting with the robot. The graph sometimes falls below 0. In particular, at the end of the experiment, it became unstable and even lower. He covered the robot’s eye camera, touched it like he was irritated, and went away from the robot. We consider that those two examples suggest the validity of the entrainment score estimation.

Evaluation of the implemented behaviors

In the sections above, we explained the analysis of body movement interaction. Here, we evaluate the implemented behaviors. Although the application of this result is limited to our approach, our findings also prove the validity and applicability of the entrainment score.

We calculated the evaluation score of each *situated module* based on the average of the entrainment score while each module was being executed. Tables 5 and 6 indicate the worst and best five modules, respectively, and their scores. The worst modules were not so interactive. SLEEP_POSE and FULLY_FED do not respond to human action and exhibit behavior similar to the sleeping pose. NOT_TURN is the behavior for brushing off a human’s hand while saying “I’m busy” when someone touches on its shoulder. The best modules were rather interactive modules that entrain humans into the interaction. EXERCISE and CONDUCTOR produce the exercising and imitating of musical conductor behaviors, which induced human synchronized body movements. Other highly rated modules also produce attractive behaviors, such asking and calling, which induce human reactions. We believe that the entrainment scores provide plenty of information for developing interactive behaviors of robots that communicate with humans.

	Coefficient	Value
Distance	α_{dist}	0.173
Eye contact	α_{ec}	0.476
Eye height	α_{eh}	0.019
Distance moved	α_{dm}	-0.228
Distance moved by hands	α_{dmh}	-0.029
Synchronized movements	α_{sm}	0.535
Touch	α_{touch}	-0.186

Table 4: Standardized partial regression coefficients obtained by multiple linear regression analysis

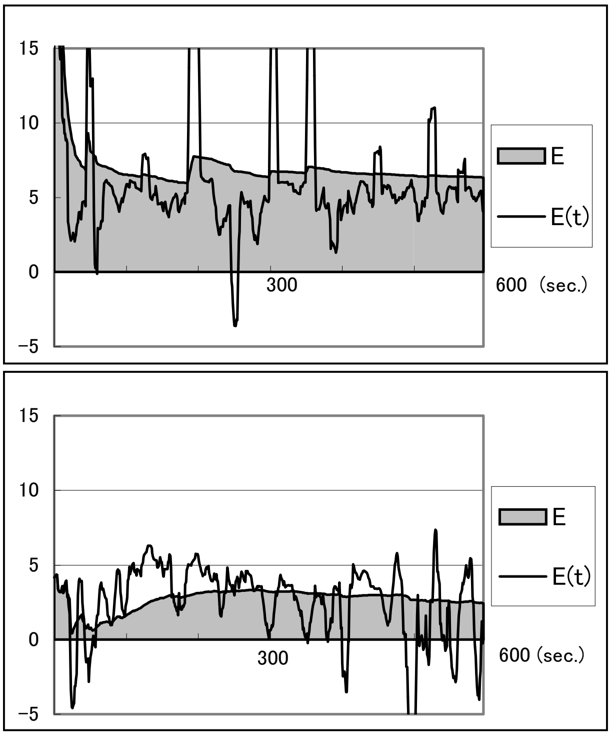

Figure 6: Illustration of entrainment score
(upper: subject who treated the robot as if it were a humans child, lower: subject who was embarrassed by interacting with it)

4 Discussions

The experiment reveals the correlation between humans’ subjective evaluations and body movements. If a human

ID	Contents	Evaluation
TICKLE	Tickle	-2.09
APOLOGIZE	Apologize	-1.96
NOT_TURN	Say, "I'm busy," and refuse to play together	-0.51
SLEEP_POSE	A pose of sleeping	-0.42
FULLY_FED	A pose of being fully fed	0.32

Table 5: Worst 5 *situated modules* based on average entrainment score

evaluates the robot highly, then the human behaves cooperatively with it, which will further improve its evaluation. That is, once they establish cooperative relationships with the robot, they interact well with the robot and evaluate the robot highly. Regarding evaluation of the implemented behaviors, the modules that entrain humans into interaction were highly evaluated, such as asking something that induces human's answer and producing cheerful body movements like exercising to let humans join and mimic the movements. We believe that the entrainment can help us to establish cooperative relationships between humans and robots.

Meanwhile, the multiple linear regression explains 59% of the subjective evaluation. This is remarkable because it is performed without regard to the contents or context of language communication. With speech recognition, the robot can talk with humans, although its ability is similar to that of a little child. Some of the subjects spoke to the robot. Often, there were requests for the robot to present particular behaviors (especially behaviors it had performed just previously), to which it sometimes responded correctly and sometimes incorrectly. To analyze this, we could use several analytical methods such as conversation analysis, however, these methods are rather subjective. On the other hand, our evaluation employed objective measures only: numerically obtained body movements without context, which means there could be a lot of potential usages. For example, an interactive robot could learn and adjust its behavior by using this method. It would be applicable to different subjects (age, culture, etc.), different agents (physical-virtual, body shape, behaviors, etc.), and inter-human communication.

5 Conclusions

This paper reported a new approach to analyzing embodied communication between humans and a robot. Our interactive humanoid robot is able to autonomously interact with humans. This complexity and autonomy is achieved by many simple behaviors. We measured humans' body movements while they observed and interacted with the robot, and the result of the analysis indicates positive correlations between cooperative body movements and subjective evaluations. Furthermore, the multiple linear regression explains 59% of the subjective evaluation without regard to language communication. We consider our approach of body movement analysis to be widely applicable in embodied communication.

ID	Contents	Evaluation
EXERCISE	Exercise	5.75
ASK_SING	Ask humans, "May I sing a song?"	5.59
CONDUCTOR	Pose imitating a musical conductor	4.85
WHERE_FROM	Ask humans, "Where are you from?"	4.55
LET'S_PLAY	Say, "Let's play, touch me"	4.24

Table 6: Best 5 *situated modules* based on average entrainment score

References

[Breazeal *et al.*, 1999] C. Breazeal and B. Scassellati, A context-dependent attention system for a social robot, *Proc. Int. Joint Conf. on Artificial Intelligence,* pp.1146-1151, 1999.

[Hall, 1966] E. Hall, *The Hidden Dimension*, Anchor Books/Doubleday

[Hirai *et al.*, 1998] K. Hirai, M. Hirose, Y. Haikawa, and T. Takenaka, The development of Honda humanoid robot, *Proc. IEEE Int. Conf. on Robotics and Automation,* 1998.

[Jebara and Pentland, 1999] T. Jebara and A. Pentland, Action Reaction Learning: Automatic Visual Analysis and Synthesis of Interactive Behaviour, *Int. Conf. on Computer Vision Systems,* 1999.

[Kanda *et al.*, 2002a] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu, Development and Evaluation of an Interactive Humanoid Robot "Robovie", *IEEE Int. Conf. on Robotics and Automation,* pp.1848-1855, 2002.

[Kanda *et al.*, 2002b] 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.

[Nakadai *et al.*, 2001] K. Nakadai, K. Hidai, H. Mizoguchi, H. G. Okuno, and H. Kitano, Real-Time Auditory and Visual Multiple-Object Tracking for Robots, *Proc. Int. Joint Conf. on Artificial Intelligence*, pp. 1425-1432, 2001.

[Nakauchi *et al.*, 2002] Y. Nakauchi and R. Simmons, A Social Robot that Stands in Line, *Autonomous Robots*, Vol. 12, No. 3, pp. 313 – 324, 2002.

[Ono *et al.* 2001] T. Ono, M. Imai, and H. Ishiguro, A Model of Embodied Communications with Gestures between Humans and Robots, *Proc. of Twenty-third Annual Meeting of the Cognitive Science Society*, pp. 732-737, 2001.

[Reeves and Nass, 1996] B. Reeves and C. Nass, *The Media equation,* CSLI Publications, 1996.

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