2005 IEEE/RSJ International Conference on Intelligent Robots and Systems

# Spatially Mapping of Friendliness for Human-Robot Interaction

Tsuyoshi Tasaki, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno
*Graduate School of Informatics, Kyoto University*
*Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan*
{*tasaki, komatani, ogata, okuno*}*@kuis.kyoto-u.ac.jp*

***Abstract*— It is important that robots interact with multiple people. However, most research has dealt with only interaction between one robot and one person and assumed that the distance between them does not change. This paper focuses on the spatial relationships between a robot and multiple people during interaction. Based on the distance between them, our robot selects appropriate functions to use. It does this using a method we developed for spatially mapping the "friendliness" of each space around the robot. The robot interacts with the highest friendliness spaces (people) selectively, thereby enabling interaction between the robot and multiple people. Our humanoid robot, SIG2 which the proposed method was implemented into, interacted with about 30 visitors, at the Kyoto University Museum. The results obtained using questionnaires after interaction showed that the actions of SIG2 were easy to understand even when it interacted with multiple people at the same time and that SIG2 behaved in a friendly manner.**

***Index Terms*— mapping of friendliness, interaction partner, multiple people**

## I. Introduction

Many humanoid robots have been produced, and people have many chances to interact with robots[1]. Therefore, robots must easily interact with people. To interact at an advanced level, robots must recognize the environment and focus on appropriate objects such as an interaction partner. For example, Murakita et al. [2] localized people exactly using touch sensors installed over the whole floor, which requires installing many devices. Kanda et al. [3] used ID tags to localize people, but asking everyone to wear a tag is inconvenient. Miyashita et al. [4] enhanced the accuracy of localizing a particular person by considering other people as noise. Regarding most people as noise is problematic for social robots. We previously described a method for controlling a robot so that it changes its function based on the distance to the interaction partner and for effectively localizing people [6].

Although there has been much work related to localization, it is difficult to localize people exactly in various environments using only the robot's sensors. Moreover, little work has been reported for selecting a person as an interaction partner from multiple people.

We have now developed a method for selecting an interaction partner for a robot based on the degree of friendliness as mapped onto the "space", considering whether people exist or not. Our aim is to achieve interaction based on the robot localizing people robustly in various environments, and for the robot to impress the people interacting with it simultaneously as intelligent and friendly.

In Section II, the distance between the robot and people during an interaction is discussed. In Section III, we describe our "friendliness space map" showing how "friendliness" is distributed in the space. In Section IV, the humanoid robot used in this study and the method for selecting an interaction partner are described. In Section V, our evaluation method is described and results are presented. In Section VI, the results are discussed, and in Section VII, the paper is summarized, and future work is mentioned.

## II. Distance between Robot and People during Interaction

### A. Interaction Distance of People

When people interact with each other, the distance between them is associated with their degree of friendliness. Proxemics [5], which is a social psychology theory, says that two people interact at an appropriate physical distance from one another based on their relationship. In this theory, the interaction distance can be classified into roughly four groups: intimate, personal, social, and public.

- Intimate distance (approx. 50 cm)
  People can communicate via physical interaction and express strong emotions.
- Personal distance (approx. 50–120 cm)
  People can talk intimately.
- Social distance (approx. 120–360 cm)
  People don't know each other well.
- Public distance (approx. 360 cm and more )
  People who have no personal relationship with each other can comfortably coexist at this distance.

These distances can be used to set the degree of friendliness between the robot and each person. The distances shown in parentheses are only typical ones. They depend on each person's personality and cultural background.

### *B. Effective Distance and Advantages and Disadvantages of Robot's Functions*

Since most functions and devices used by a robot are not effective for all distances, we assessed the effective distance for them. We investigated the effective distance of tactile recognition, speech recognition, sound source localization, and face localization, which are implemented into many robots as general functions.

*1) Tactile Recognition:* Tactile recognition is done using tactile sensors, which are effective when people can touch the robot. The average length of a person's arm is about 70 cm, so the appropriate distance for tactile recognition is up to 50 cm. This distance is similar to the intimate distance.

*2) Speech Recognition:* To determine the range for speech recognition, we place a speaker in front of a robot at every 50 cm from 50 cm to 3 meters and played 200 words of the ATR phonetically balanced corpus. The results of isolated word recognition using "Julian" [7], general Japanese automatic speech recognition software, are shown in Fig. 1. Automatic speech recognition was found to be effective up to around 1.5 meters.

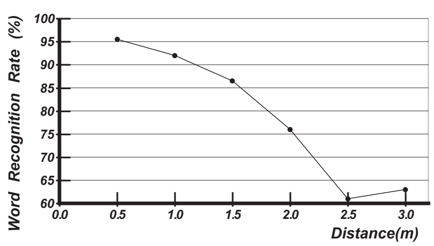


Fig. 1. Isolated Word Recognition at Various Distances

*3) Sound Source Localization:* A well-known sound source localization function uses the Interaural Phase Difference (IPD) and Interaural Intensity Difference (IID) [8]. The effective distance of sound source localization on average and the standard deviations were estimated in our laboratory (Fig. 2). Three directions were evaluated separately. The horizontal direction was specified from right (0°) to left (180°), and the center was 90°.

The localization errors were small for distances less than about 3 meters. Therefore, sound source localization should be stable up to around 3 meters.

*4) Face Localization:* We use MPIsearch [9] for robust face detection. A robot can measure the distance and direction to a person based on the average size of a person's face.

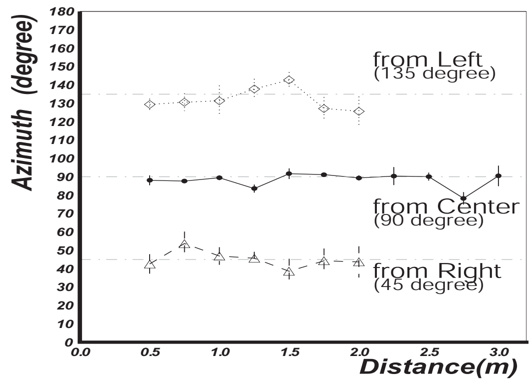


Fig. 2. Sound Source Localization at Various Distances

MPIsearch requires an image at least 12 by 12 pixels to detect a face. Such images correspond to a distance of 4 to 5 meters. In general, the effective distance of face localization is up to the public distance.

*5) Advantages and Disadvantages of Functions:* The advantages and disadvantages of tactile recognition, sound source localization, and face localization are shown in Table. I. While tactile recognition can localize a person within the length of a person's arm, it cannot detect the direction to the person precisely. While sound source localization can detect the direction to the person exists and is not affected by occlusion, it is affected by environmental sound (noise). While face localization can detect not only the distance but also the direction to the person, it suffers if the lighting is poor.

Considering these factors, the integration of several functions into a robot enables a robot to localize people more robustly. For example, if poor lighting impairs face localization, tactile recognition and sound source localization can be used instead.

TABLE I

ADVANTAGES AND DISADVANTAGES OF ROBOT FUNCTIONS

| Function | Advantages | Disadvantages |
|---|---|---|
| Tactile Recognition | near distance detection<br>high reliability | weak direction detection |
| Sound Source Localization | direction detection<br>no occlusion effects | mixed sound effects |
| Face Localization | direction detection<br>distance detection | light effects |

### *C. Interaction Distance and Effective Distance of Functions*

The relationship between the interaction distance and the effective distance for the three functions is shown in Table II. As shown in the Table II, effective distance for the functions can correspond to the interaction distance effectively.

TABLE II

RELATIONSHIP BETWEEN DISTANCE AND FUNCTION

| Intimate Distance | Personal Distance | Social Distance |
|---|---|---|
| Tactile Recognition | | |
| Speech Recognition | Speech Recognition | |
| Face Localization | Face Localization | Face Localization |
| Sound Localization | Sound Localization | Sound Localization |

## III. FRIENDLINESS SPACE MAP

### A. Design Friendliness Space Map

In various environments, the sensor inputs capture noise. Moreover, the sensor functions a robot can use effectively differ depending on the distance between the robot and each person.

In other relational studies, the robot always used all sensors and interacted with people by focusing on the people. In our study, the robot interacted with people by focusing on the "space" of the people. In particular, the robot acted based on the space around the robot, segmented as described in Table II.

Given the size of a person's face and the accuracy of the robot's functions, the direction element of space must be segmented to some extent. We segmented the space every 15 degrees based on the average size of the human face (16 cm × 23 cm) and the errors of functions within the personal distance.

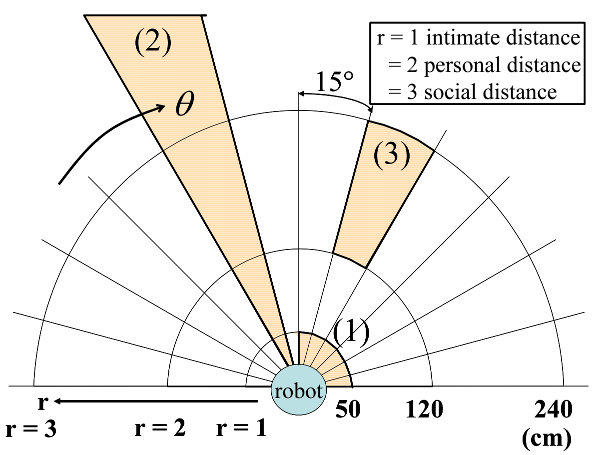


Fig. 3. Friendliness Space Map and Effective Area of Functions

To identify the intimate space for the robot to interact with, we defined polar coordinates as shown in Fig. 3. These coordinates, which are segmented into cells, are called a "Friendliness Space Map". Our robot calculates the "friendliness" of a cell (r, $\theta$) using information about the location of people and comfortable/uncomfortable stimuli. To calculate the friendliness, when a function is initiated by sensor input, our robot calculates the Human Existence Degree (HED), which shows whether people exist or not, of cells within the effective area of each function. For example, three areas where our robot calculated the HED are shown in Fig. 3: (1) in the case the right side of the robot is touched, (2) in the case the robot detects sound, (3) in the case the robot detects face.

The effects of interaction using this map are as follows.

- Since a robot can change its motion and select an interaction partner based on the friendliness of various spaces, it can interact with multiple people simultaneously in various environments.
- The action selection based on space can also be applied to various other objects.

### B. Definition of Human Existence Degree by Integration of Functions

In each cell on the map, the HED is calculated by taking advantage of the integrated functions. When a function $k$ locates a person at time $t_{k0}$, it calculates the HED, $L_{k,t,r,\theta}$, of cell (r, $\theta$) within the effective function area at time $t$, as shown Eq. (1). The $k$ ($k$ = 1, 2, 3) is the functions, $d_k$ is the damping ratio (which is decided based on the degree of confidence obtained by previous experiments of each function) , and $t_{k0}$ is renewed every time function $k$ operates.

$$L_{k,t,r,\theta} = \exp[-d_k(t - t_{k0})] \tag{1}$$

The HED calculated by integration of all functions, $E_{t,r,\theta}$, of cell (r, $\theta$) at time $t$ is defined as the sum of the HED of each function:

$$E_{t,r,\theta} = \sum_{k=1}^{3} L_{k,t,r,\theta}. \tag{2}$$

### C. Shift in Friendliness by Stimulus

The cells on the Friendliness Space Map are affected by the kind of stimulus. Our robot recognizes two kinds of stimuli by using tactile recognition. One is uncomfortable stimuli, such as hitting the robot's head or touching the robot's bust. The other is comfortable stimuli, such as patting the robot's head. Since tactile recognition cannot localize people precisely, we assume the person delivering the stimulus is in the cell with the highest human existence degree within the intimate distance. That is, it is cell (1, $\theta$), as obtained using

$$\theta = argmax_{\theta} E_{t,1,\theta}. \tag{3}$$

If the stimulus occurs at time $t_{C0}$, we define the Comfortable Degree (CD), $C_{t,1,\theta}$, of cell (1, $\theta$) selected at time $t$ as shown in Eq. (4), where $d_C$ is the damping ratio, $v$ is the kind of stimulus ($v$ 1 is a comfortable stimulus and $v$ -1 is an uncomfortable stimulus), and $t_{C0}$ is renewed every time a stimulus is received.

$$C_{t,r,\theta} = v \times \exp[-d_C(t - t_{C0})] \quad (4)$$

### D. Definition of Friendliness

The Friendliness Space Map is renewed and consists of both the HED and the CD obtained using the robot's functions. The friendliness, $I_{t,r,\theta}$, of cell (r, $\theta$) at time t is defined as the sum of the HED and the CD as shown in Eq. (5), where $W_L$ and $W_C$ correspond to the weights of the HED and the CD, respectively. In this time, we make $W_C$ bigger than $W_L$ because we want a robot to be sensitive to the stimulus.

$$I_{t,r,\theta} = W_L \times E_{t,r,\theta} + W_C \times C_{t,r,\theta} \quad (5)$$

## IV. Human-Robot Interaction Based on Friendliness Space Map

### A. Humanoid Robot SIG2

The platform we used is the humanoid robot SIG2 shown in Fig. 4 (left). It has 19 tactile sensors on its head and upper body, a microphone ("ear") on each side of its head, and two cameras ("eyes") in its head. To improve reception, each microphone is embedded at the eardrum of a human outer ear model made of silicon, as shown in Fig. 4 (right). SIG2 utters and gestures by using a speaker and three motors in its head.

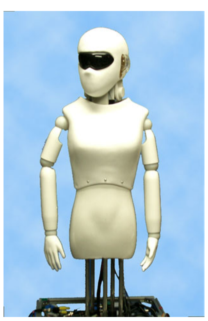

Fig. 4. SIG2 (left) and One Ear (right)

SIG2 is equipped with tactile recognition, speech recognition, sound source localization, and face localization. The tactile recognition recognizes the spot on the robot touched by a person and two kinds of contact (hitting and patting) using the touch duration. The speech recognition recognizes the numbers 1–15, and "yes", and "no". The sound source localization and face localization was showed in Section II-B. The output functions are tactile reaction, game dialogue, intimate person selection, trace face and trace sound.

- Tactile Reaction
  SIG2 can perform five types of actions such as a delighted action or a sad action based on both the spot touched and the kind of stimulus.
- Game Dialogue
  SIG2 can play a game using speech recognition if there is an intimate person within the personal distance by using speech recognition. In this game, SIG2 and its interaction partner say random numbers from 1 to 15 to each other. They can repeat the number up to four times at once. The first one who says a number that has already been said loses. SIG2 uses gestures and utterances that match the situation of the game.
- Intimate Person Selection
  After gesturing and uttering using other output functions, SIG2 turns on the cell direction that has the highest friendliness level within the personal distance using this function.
- Face Trace and Sound Trace SIG2 gazes at the direction where it finds a person's face or hears a sound.

### B. Design of Interaction Using Friendliness Space Map

More specifically, SIG2 interacts with people as follows.

1) SIG2 turns on the direction calculated by tactile recognition, face localization, or sound source localization. If SIG2 uses tactile recognition, it acts based on both the kind of stimulus and the spot touched in accordance with the outputs of the tactile recognition.
2) After referring to the friendliness space map, SIG2 renews it based on the results of person localization and stimulus type.
3) If the stimulus is comfortable and the friendliness of the cell within the personal distance exceeds a threshold, SIG2 plays a game with the person in that cell.
4) SIG2 turns on the direction of the cell that has the highest friendliness level on the friendliness space map.

## V. Evaluation

### A. Effectiveness of the Person Localization

*1) Aim and Sequence of Experiment:* To determine whether a person is in the direction where SIG2 feels intimate in an actual environment, we compared the accuracy of sound source localization, which is the most accurate of the three functions, with the accuracy of proposed method, at Kyoto University Museum. Testing was done during the daytime, so the museum was illuminated by both natural and artificial light. Moreover, museum announcements were broadcasted regularly. Testing was done using seven pairs of participants.

1) We explained to the participants the input functions of SIG2.
2) SIG2 interacted with each pair for about 5 minutes.

The evaluation criteria were the recall ratio, precision ratio, and F value. They were calculated when the detected direction corresponded with one of the people, during their interaction, and when the system detected people.

*2) Results:* The relationship between the distribution of cells which had the highest friendliness level at the intimate distance and the directions in which there were people is shown in Fig. 5. Two people interacted with SIG2 at cells (1, 6) and (1, 9) which correspond to person on left and person on right in Fig. 5 respectively. In Fig. 5, we can see that there were people in the cell with the highest friendliness level.

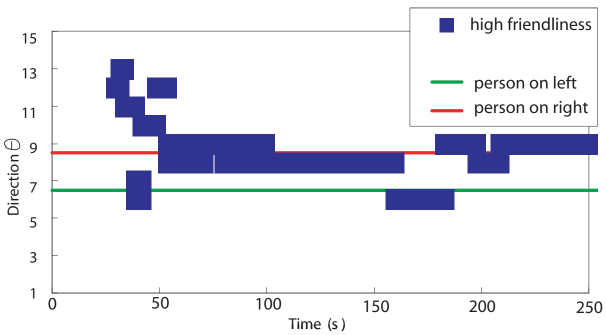


Fig. 5. Relationship Between Friendliness Distribution and People Direction

The recall ratio, precision ratio, and F value are shown in Table III. The F value with the proposed method was higher than with only sound source localization.

TABLE III
ACCURACY OF PERSON LOCALIZATION

| | Only Sound Localization | Friendliness Space Map |
|---|---|---|
| Recall | 0.52 | 0.83 |
| Precision | 0.33 | 0.71 |
| F Value | 0.40 | 0.76 |

### *B. Impression Evaluation of Interaction using Friendliness Space Map*

*1) Aim and Sequence of Experiment:* We investigated whether our method enabled SIG2 to make a plausible and friendly impression when interacting with several people simultaneously. We asked 27 visitors (men and women ranging in age from 20 to 54) to interact with SIG2 at Kyoto University Museum and then fill out a questionnaire. The experimental conditions were the same as described in Section V-A. The experimental setup is shown in Fig. 6.

TABLE IV
EVALUATED ADJECTIVE PAIRS AND FACTOR MATRIX

| Adjective pairs | | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|---|
| Kind | Cruel | -0.103 | **0.720** | 0.149 | -0.110 |
| Favorable | Unfavorable | 0.094 | **0.689** | -0.110 | -0.186 |
| Friendly | Unfriendly | 0.315 | **0.681** | -0.061 | -0.028 |
| Safe | Dangerous | 0.204 | 0.517 | -0.191 | 0.257 |
| Warm | Cold | 0.275 | 0.550 | 0.262 | -0.043 |
| Pretty | Ugly | 0.220 | **0.661** | 0.195 | 0.011 |
| Frank | Rigid | 0.535 | 0.291 | 0.143 | -0.004 |
| Distinct | Vague | **0.636** | -0.099 | -0.092 | -0.474 |
| Accessible | Inaccessible | 0.522 | 0.265 | 0.061 | -0.072 |
| Light | Dark | 0.470 | 0.318 | 0.329 | 0.051 |
| Altruistic | Selfish | 0.260 | 0.155 | 0.188 | -0.493 |
| Humanlike | Mechanical | 0.413 | 0.289 | 0.186 | -0.035 |
| Full | Empty | **0.604** | -0.027 | 0.058 | 0.119 |
| Exciting | Dull | **0.857** | 0.002 | -0.196 | 0.023 |
| Pleasant | Unpleasant | **0.805** | 0.138 | -0.159 | 0.101 |
| Likable | Dislikeable | **0.857** | 0.122 | -0.034 | -0.137 |
| Interesting | Boring | 0.497 | 0.442 | -0.246 | 0.027 |
| Good | Bad | **0.734** | 0.151 | -0.183 | 0.004 |
| Complex | Simple | 0.045 | 0.058 | 0.419 | 0.139 |
| Rapid | Slow | -0.103 | 0.007 | **0.910** | -0.153 |
| Quick | Slow | -0.147 | 0.017 | **0.808** | -0.109 |
| Agitated | Calm | -0.020 | -0.499 | 0.484 | 0.109 |
| Active | Passive | 0.105 | 0.108 | 0.493 | 0.498 |
| Brave | Cowardly | 0.076 | -0.136 | 0.076 | **0.761** |
| Showy | Quiet | **0.674** | -0.448 | 0.378 | 0.110 |
| Cheerful | Lonely | 0.401 | 0.311 | 0.350 | 0.135 |
| Sharp | Blunt | 0.009 | 0.149 | 0.552 | 0.058 |
| Intelligent | Unintelligent | **0.801** | -0.010 | -0.014 | -0.244 |

Each groups of visitors interacted with SIG2 two times, and SIG2 used a different behavior each time. One time it behaved based on the friendliness space map, as described in Section IV-B. The other time it did not use friendliness space map to isolate the effects of our method. In the latter, SIG2 turned in the direction calculated by three functions and played the game regardless of the friendliness level if someone was within the personal distance. SIG2 selected which behavior to use at the beginning randomly. Each group interacted with SIG2 for about 5 minutes each time. Then, they filled in a questionnaire, rating 28 adjective pairs (in Japanese) on 1-to-7 scales, where 7 means the positive adjectives fit very well (adjectives in the leftmost column in Table IV), based on the SD method. This evaluation method is based on “Psychological analysis on human-robot interaction” [10].

*2) Results:* Factor analysis was performed on the SD method ratings for the 28 adjective pairs using the results of 54 (27×2) questionnaires. The factor matrix, with the factor loadings, is shown in Table IV. Referring to the adjective pairs that have loadings greater than 0.6, the first factor contains “Distinct”, “Exciting”, and so on, and the second factor contains “Kind”, “Friendly”, and so on. The first and second factors are similar to the ones obtained by Kanda et

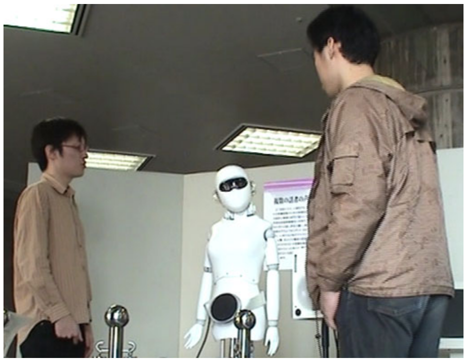

Fig. 6. Experimental Setup

al. [10]. Therefore, we think the two types of behaviors of SIG2 can be compared meaningfully using first and second factors.

Table V shows the average and standard deviations of the impression scores for the two types of behaviors for the first and second factors. T-verification showed that the difference between the two types was significant at the 0.05 level, indicating that the behavior based on the friendliness space map was considered more positive adjectives.

TABLE V
COMPARISON OF IMPRESSION SCORES

| | First Factor | | Second Factor | |
|---|---|---|---|---|
| Type | Average | S.D. | Average | S.D. |
| Based on Map | 4.59 | 1.58 | 4.79 | 0.97 |
| Ignore Map | 4.35 | 1.75 | 4.65 | 0.85 |

## VI. DISCUSSION

### A. Person Localization Based on Friendliness Space Map

We can verify that someone is in the space with the highest friendliness level, since the friendliness space map considers the human existence degree. However, the results presented in Section V-A showed that the recall ratio was low. This is because the people did not use SIG2's functions positively to interact between the person and the person ignoring SIG2. This is a special problem for interaction between a robot and "multiple" people. Therefore, we have to develop the method which enables a robot to join the interaction between multiple people appropriately.

### B. Impression of Behavior Based on Friendliness Space Map

For the times when SIG2 behaved based on the friendliness space map, the impression scores of the adjectives related to the first factor were high. This is because the simple selection criteria based on the friendliness made the SIG2 behaviors seem plausible. For the times when did not use the map, the simple behaviors resulted in lower impression scores.

If robots can behave richly and plausibly even when interacting with multiple people, they might be a member of the interaction group.

## VII. CONCLUSION

We have developed a human-robot interaction method based on the "friendliness space map", which focuses on the "space" rather than the person to find and select interaction partners in various environments. An experiment done at Kyoto University Museum showed that this method enabled the SIG2 to locate and select interaction partners. Moreover, the results obtained using a questionnaire showed that SIG2 interacted with visitors in a plausible and friendly manner.

With this method, the behavior of SIG2 with interaction partners is simple. Therefore, if multiple people interact with SIG2 more than a few minutes, the person-to-person interactions increase, and SIG2 loses its impression scores. For more active interaction, the robots must interact appropriately to impress the people. We plan to implement the proposed method in a robot that has many degrees of freedom and behaves using Q-learning with friendliness as a reward.

## ACKNOWLEDGMENTS

This research was supported by the JPSP 21st Century COE program on informatics research for development of knowledge society infrastructure and SCAT.

## REFERENCES

[1] Breazeal. C.L.: "Designing Sociable Robots", A Bradford Book, ISBN 0262025108, 2001.

[2] T. Murakita, T. Ikeda, and H. Ishiguro, "Human Tracking using Floor Sensors based on the Markov Chain Monte Carlo Method", Seventeenth International Conference on Pattern Recognition (ICPR), pp. 917-920, Aug. 2004.

[3] T. Kanda and H. Ishiguro, "Reading Human Relationships from Their Interaction with an Interactive Humanoid Robot", International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE), pp. 402-412, May 2004.

[4] T. Miyashita, M. Shiomi, and H. Ishiguro, "Multisensor-based Human Tracking Behaviors with Markov Chain Monte Carlo Methods", Proceedings of IEEE-RAS/RSJ International Conference on Humanoid Robots, Nov. 2004.

[5] Hall. E. T., "Hidden Dimension. Doubleday Publishing", ISBN 0385084765, 1996.

[6] T. Tasaki, S. Matsumoto, H. Ohba, M. Toda, K. Komatani, T. Ogata, and H. G. Okuno: "Dynamic Communication of Humanoid Robot with Multiple People Based on Interaction Distance", in Proc. of IEEE International Workshop on Robot and Human Interaction (ROMAN 2004), pp. 71-76, 2004.

[7] http://julius.sourceforge.jp/

[8] H. G. Okuno, K. Nakadai, K. Hidai, H. Mizoguchi, H. Kitano, "Human-Robot Interaction Through Real-Time Auditory and Visual Multiple-Talker Tracking", IROS, 2004.

[9] Fasel, I. and Movellan, J.R., "Comparison of neurally inspired face detection algorithms", UAM, 2002. Proc. of International Conference on Artificial Neural Networks (ICANN 2002), 1395-1401, 2002.

[10] T. Kanda, H. Ishiguro, and T. Ishida, "Psychological analysis on human-robot interaction", IEEE International Conference on Robotics and Automation (ICRA 2001), pp. 4166-4173, 2001.