KANSEI EVALUATION OF BEHAVIORS OF ROBOT WHICH RECOGNIZES DIFFERENCE BETWEEN USER'S AND ITS OWN FIELDS OF VIEW

Jun-ichi IMAI[*a] and Masahide KANEKO[a]

[a] *The University of Electro-Communications, Japan.*

ABSTRACT

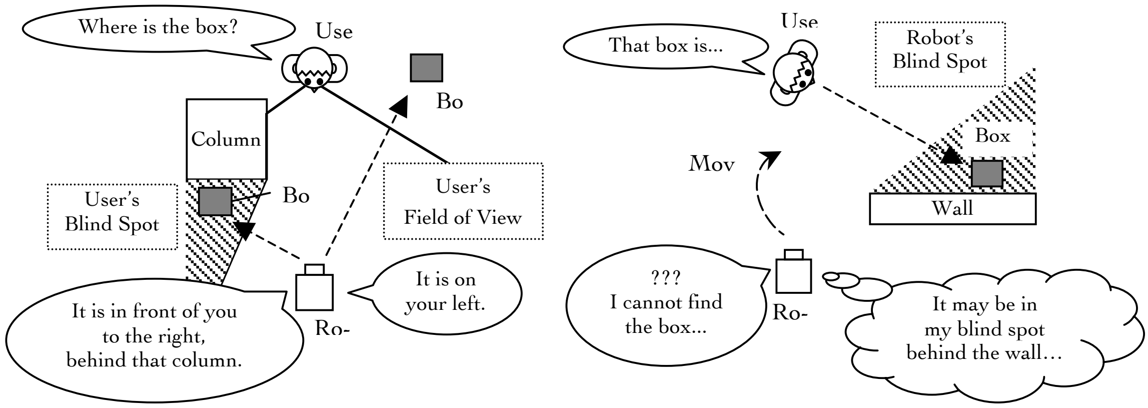
Generally there are many objects which can cause visual occlusion in daily living spaces for humans, where human-symbiotic robots will work. Consequently, it will often occur that a robot cannot see an object by occlusion while a user can, and vice versa. In such situations, it is desirable for the robot to be able to interact with a user while recognizing a difference between their fields of view. We expect that such a "considerate" robot will be friendlier and more pleasant to users. In this paper, we carry out experimental subjective evaluations of impressions which such robot gives to humans during human-robot interactions to verify our expectation. We have developed a robot which can estimate a user's and its own fields of view to behave appropriately while recognizing the difference between their perceptions in our previous works. Participants are requested to observe the interactions in occlusion environments and to subjectively evaluate impressions which they receive from the developed robot's behaviors. The experimental results show that the robot which can guess a user's perception and understand differences between their recognition of situations can give "familiar" impressions to humans. This fact is expected to be one of fundamental recommendations for designing much friendlier interactions with robots and other intelligent systems.

Keywords: *Human-robot interaction, Impression evaluation, Visual occlusion, View estimation, Difference in perceptions*

1. INTRODUCTION

Human-symbiotic robots, which can communicate with humans and support their activities, have been increasingly studied in recent years (e.g. [1–3]). They will work mainly in daily living spaces for humans, such as offices, homes and public spaces. In these environments, generally there are many objects which can cause visual occlusion. Since a robot and a

* **Corresponding author:** 1–5–1 Chofugaoka, Chofu-shi, Tokyo 182–8585, Japan.



(a) The case that only a robot can see boxes. (b) The case that only a user can see a box.

Figure 1: Examples of interactions between a robot and a user in occlusion environments.

user will stand at different positions, it will often occur that one cannot see an object by occlusion while the other can. In such situations, it is desirable for the human-symbiotic robot to be able to interact with a user while considering a difference between their fields of view and putting itself in his or her position. Examples of such interactions are shown in Figure 1. In Figure 1 (a), a robot is expected to recognize that boxes are out of a user's view and to prompt the user to find the boxes by giving appropriate directions. In Figure 1 (b), a robot is expected to understand that a box cannot be seen from its own position and to move to the appropriate position autonomously. We expect that such a "considerate" robot will be friendlier and more pleasant to users. However, there have been few studies which deal with these problems explicitly so far. The aim of this paper is to verify this expectation.

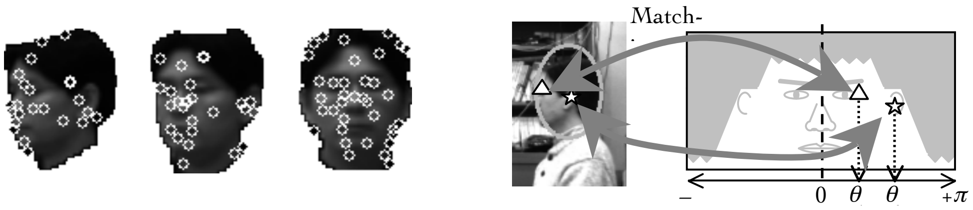
In our previous works, we have developed a robot which can estimate a user's and its own fields of view to behave appropriately while recognizing the difference between their perceptions [4, 5]. In this paper, using this developed robot, we carry out experimental subjective evaluations of impressions which the robot gives to humans during human-robot interactions. The robot performs several tasks with one user in real occlusion environments. Participants who observe the interactions are requested to subjectively evaluate impressions which they get from the robot's behaviors by questionnaires. Based on their evaluation results, we discuss the effectiveness of the robot which can recognize the difference in fields of view.

2. ROBOT WHICH CAN ESTIMATE USER'S AND ITS OWN FIELDS OF VIEW

In this section, we briefly review the robot which can estimate a user's and its own fields of view developed in our previous works [4, 5].

2.1. Estimation of User's State

To estimate a user's view, the user's current state is firstly estimated from an image obtained by a stereo camera. The estimation is achieved by two steps. First the user's head position is detected, and then its orientation is estimated.



(a) Extracted SIFT features. (b) Matching of feature points.

Figure 2: An example of estimation of user's state.

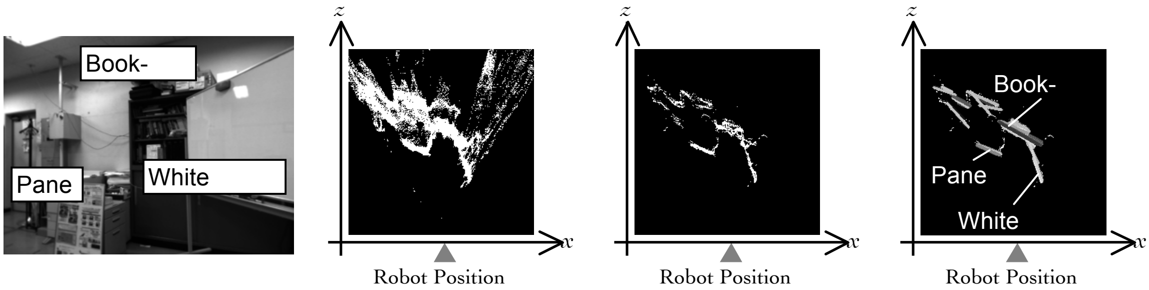
We use the particle filter [6] for tracking a position of the user's head. The user's head in the input image is modeled by an ellipse. The position of the ellipse for the user's head is detected using skin color information [7], hair color information and distance information (obtained by the stereo camera) in and around the ellipse. Based on the specified head position, the orientation of the user's head is estimated. We use a head model using the SIFT features [8] for the estimation. In this method, the user's head is modeled by a cylinder. The center of the user's face (the position of the user's nose) is defined as 0 [rad] and the back of the head is defined as $\pm\pi$ [rad]. SIFT feature points extracted from the specified head image are stored with their angle positions measured from the center of the face. Figure 2 shows an example of estimation of the user's head orientation. Figure 2 (a) shows an example of extracted feature points from the specified head image. We can estimate the user's head orientation by performing matching procedure between the extracted feature points and stored ones in the head model. Figure 2 (b) shows an example of this process. The ellipse in Figure 2 (b) denotes the specified position of the user's head. The angle position of the center of the head image measured from the center of the face can be calculated using values of $\theta 1$, $\theta 2$, ... and the positions in the head image of their corresponding feature points.

This method can estimate the user's head orientation even in the case that the user does not show his or her frontal face to the camera. The head model can be created and updated in the online fashion without any prior learning if tracking and estimation is started in the state that the user faces to the camera.

Furthermore, the orientation of the user's body is estimated. The 3D positional data under the user's head (i.e. the user's body) obtained by the stereo camera is projected onto the quantized x-z plane. The user's body on the x-z plane is modeled by a slant ellipse. We can determine the body orientation as the slant of the ellipse obtained by fitting procedure to the projected positional data. Although the body orientation is not necessary for the view estimation, it is used in the experiments described in Section 3.

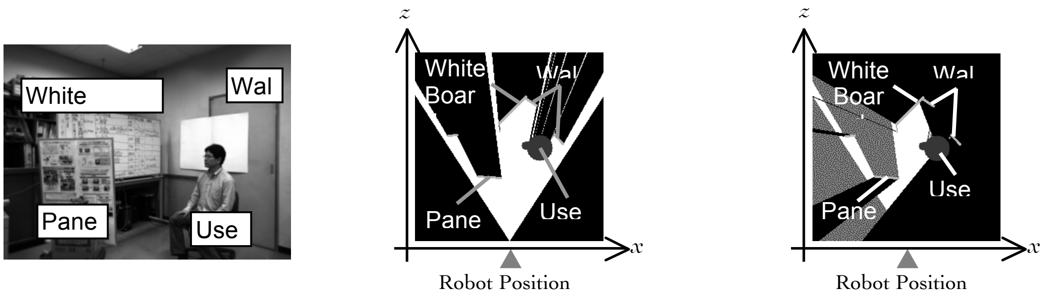
2.2. Environmental Recognition

Another main part of the view estimation is the environmental recognition. Accurate recognition of surrounding environments, which can cause occlusion, is required to estimate the user's and the robot's fields of view. However, it is difficult to recognize all objects which exist in the surrounding. So the robot attempts to extract vertical planes in its surrounding instead. This approach is valid because vertical planes are very common in artificial environments and it is highly probable that they are major causes of occlusion.



(a) Original image. (b) x-z plot. (c) After binarization. (d) Result.

Figure 3: An example of environmental recognition.



(a) Original image. (b) Estimated robot's view. (c) Estimated user's view.

Figure 4: Examples of estimated fields of view.

We use the probabilistic Hough transform [9] to detect vertical planes for the environmental recognition. Figure 3 shows an example of the recognition. Figure 3 (a) is an original image. First, all 3D positional data obtained by the stereo camera is projected onto the quantized x-z plane (Figure 3 (b)). The vertical axis of Figure 3 (b) denotes the depth from the camera. The center of x-axis corresponds to the position where the robot stands. Then, the x-z plot is binarized to remove pixels except for those corresponding to vertical planes (Figure 3 (c)). Finally line segments on the resultant x-z plane are detected using the probabilistic Hough transform. Detected line segments are shown in Figure 3 (d). These line segments correspond to the vertical planes such as a panel, a white board and a bookshelf.

2.3. Estimation of Horizontal Fileds of View

The robot's and user's horizontal fields of view are estimated by combining the results of the estimation of the user's head orientation and the environmental recognition.

The robot's view and blind spots are estimated from the positional relation between the estimated surrounding environment and the robot itself. Similarly, the user's view is estimated from the positional relation between the estimated surrounding environment and the user's position, and the user's head orientation. In this paper, we suppose that the view angle of human is $2\pi / 3$ [rad] (120°). That is, objects in the outside of the range $[\theta - \pi / 3, \theta + \pi / 3]$ are out of the user's view.

Figure 4 shows examples of estimated fields of view. In Figure 4 (a), there are a panel, a white board and walls around the user. Figure 4 (b) and (c) show the estimated robot's and user's horizontal fields of view respectively. They are the bird's-eye images viewed from just above. Gray line segments represent the surrounding environment of the user at the height of the center of his head. The white region in Figure 4 (b) represents estimated robot's own view. Therefore, the robot can understand that it cannot see objects in the black regions such as behind the panel from its position. The white and gray regions in Figure 4 (c) represent estimated user's view. It is required to distinguish the regions in which the user's view and robot's one overlap from the regions in which they do not. Overlapped regions represent the regions which the robot can visually confirm that they are within the user's view. Conversely, the other user's view represents the regions which may be within the user's view, but the robot cannot confirm whether these regions are truly open spaces or not because they are out of the robot's view. The white regions in Figure 4 (c) represent the former view, and the gray regions represent the latter one.

3. SUBJECTIVE IMPRESSION EVALUATION

3.1. Settings

We carry out a subjective evaluation of impressions given by behaviors of the robot which estimates fields of view and acts with consideration for "difference in perceptions" in human-robot interactions.

We set three tasks of interactions between a robot and a user. For each task, we set three conditions A, B and C of the robot's behaviors. Condition A is "a robot which behaves regardless of the user's state," Condition B is "a robot which behaves according to the user's state, but does not estimate fields of view (that is, does not understand 'difference in perceptions')" and Condition C is "a robot which estimates fields of view to understand the 'difference in perceptions' and behave considerate actions toward the user."

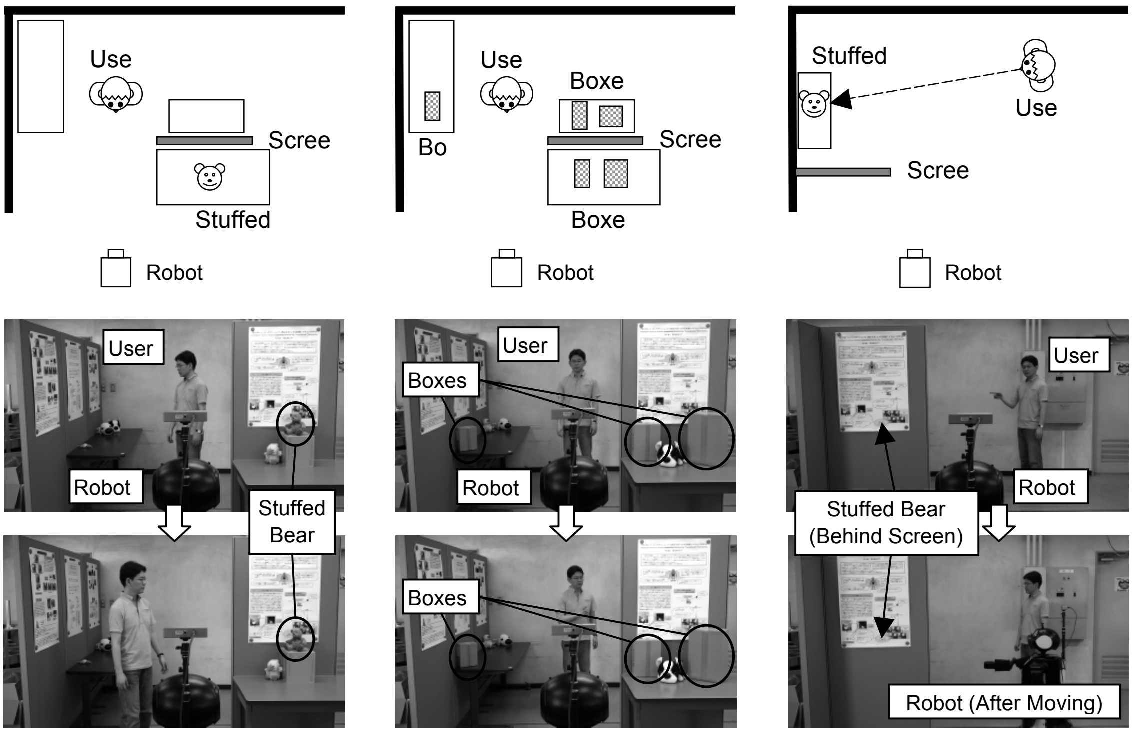
We use "Robovie-R ver.2" manufactured by Vstone Co., Ltd. as a robot in the experiments. We also use the "Bumblebee2" manufactured by Point Grey Research, Inc. as the stereo camera equipped on the robot to capture images. The image size is 320 × 240 pixels. The frame rate of our system is about 8 [fps] without any special optimization on a normal PC (Intel Core 2 Duo, 2.60 [GHz]). Details of settings for each task are as follows. Figure 5 shows bird's eye images and actual snapshots of these tasks.

[Task #1] The robot gives advices at regular intervals to a user who is looking for a stuffed bear in his blind, and prompts him to find it (Figure 5 (a)). This interaction is the almost same as that shown in Figure 1 (a).

Condition A: The robot tells the user the direction of the stuffed bear viewed from itself, regardless of the user's state. ("It is in front to right.")

Condition B: The robot tells the user the relative direction of the stuffed bear to the user's body. It does not estimate their fields of view. ("It is on the left. (If the user turns left) It is in front.")

Condition C: In addition to Condition B, when the stuffed bear is in a blind spot in front of the user, the robot tells him so and beckons him. ("It is in front. You cannot see it from your position, so please come here.")



(a) Task #1. (b) Task #2. (c) Task #3.

Figure 5: Bird's eye images and snapshots of three tasks.

[Task #2] The robot and the user count cardboard boxes around them in cooperation with each other. The robot provides information to the user when he asks the robot (Figure 5 (b)).

Condition A: The robot tells the user the number of boxes which it can see, regardless of the user's state. ("I can see three boxes in all.")

Condition B: The robot tells the user the number of boxes which it can see with the relative direction of them to the user's body. ("One in front to the right, two in front to the left, there are three boxes in all.")

Condition C: The robot tells the user the number of boxes with the relative direction to the user's body for boxes which the user can see, and tells him that they are in his blind for those which he cannot see. ("One in front to the right, two in your blind, there are three boxes in all.")

[Task #3] The user turns his body to a stuffed bear in the robot's blind and instruct the robot to carry it. The robot reacts to that (Figure 5 (c)). This interaction is the almost same as that shown in Figure 1 (b).

Condition A: The robot searches the stuffed bear in the input image without gestures, regardless of the user's state. After that, it looks to the user and tells him that it could not find the bear. ("A stuffed bear could not be found anywhere.")

Condition B: The robot searches the stuffed bear along the frontal direction of the user's body while turning its head. After that, it tells the user that it could not find the bear while pointing the frontal region of his body. ("A stuffed bear could not be found in front of you.")

Table 1: Adjective pairs.

Adjective Pairs			
kind – unkind	agreeable – disagreeable	active – passive	wise – foolish
pleasant – unpleasant	human-like – mechanical	cheerful – gloomy	violent – mild
interesting – boring	enjoyable – unenjoyable	pretty – hateful	good – bad
safe – dangerous	confident – fainthearted	warm – cold	complex – simple
favorite – unfavorite	intelligible – unintelligible	full – empty	bright – dark
friendly – unfriendly	approachable – unapporoachable	quick – sluggish	fast – slow
sensitive – insensitive	considerate – selfish	showy – plain	informal – formal

Condition C: The robot confirms that there is its blind spot on the frontal direction of the user's body, and move to the appropriate position where it can see that spot. After moving, it points out the stuffed bear. ("I cannot see that from here, so I will move there. (After moving) That stuffed bear, isn't it?")

Two kinds of video of the human-robot interaction are made for each task. One is taken by a fixed camera behind the robot, and another is taken by a camera held by the user who actually interacts with the robot. Twenty males and females, aged from twenties to fifties, participate in the evaluation. Participants watch these videos and subjectively evaluate impressions which they receive from the robot's behaviors by questionnaires based on the semantic differential (SD) method [10], which includes 28 adjective pairs. The adjective pairs used in the experiment are shown in Table 1. Participants are requested to evaluate impressions concerning each adjective pair in seven grades. Score is higher if the degree of positive adjective is higher. They are also requested to answer the following two questions in seven grades.

Q1: "Did the robot behave appropriately according to the situation?"

Q2: "Did the robot behave while putting itself in the user's position?"

The order of presentation of three conditions A, B and C in each task and that of adjective pairs in the questionnaire are counterbalanced. The arrangement of objects in the experimental environment of each task (shown in the top row in Figure 5) is explained to the participants before they watch the videos. Consequently, the participants can know what the user and the robot can and cannot see when they evaluate the robot's behaviors.

The factor analysis was applied to the evaluation results for each task independently. Then five factors for Task #1, five factors for Task #2, and three factors for Task #3 were extracted respectively. The numbers of factors were determined by the eigenvalues of the correlation coefficient matrix of the evaluation score for the adjective pairs. Cumulative contributions are 0.672, 0.728 and 0.734 respectively.

3.2. Results

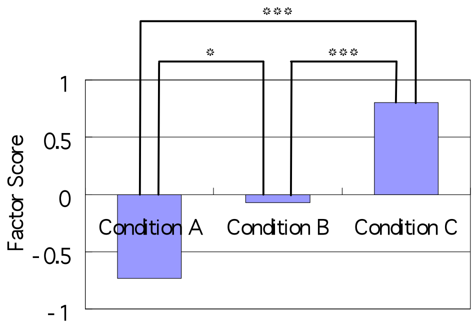
We compare impressions among three conditions in each task based on the factor scores. The analysis of variance (ANOVA) indicates that there are significant differences in the scores of several factors. Especially, we focus on the third factor of Task #1, the third factor of Task #2, and the first factor of Task #3. Table 2 shows loadings of these factors. These factors have high loadings especially for "considerate" (All tasks), "agreeable" (Task #2, #3), "approachable" (Task #1, #3) and so on. Consequently, we can interpret them as factors strongly related to good impressions on the robot, especially "familiarity." Figure 6 shows the graph of means of these "familiarity" scores. The multiple comparisons between the scores of these factors indicates that the score of Condition C is significantly higher than those of other two conditions in every task. ***, ** and * in Figure 6 represent that there are significant differences with $p < 0.001$, 0.01 and 0.05 respectively.

Table 2: Factor loadings.

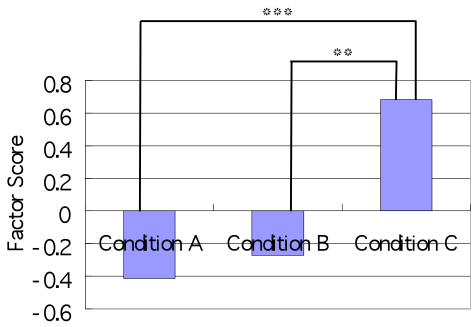
Adjective	Task #1	Task #2	Task #3	Adjective	Task #1	Task #2	Task #3
kind	0.327	0.303	**0.549**	active	0.140	0.096	**0.831**
pleasant	0.340	0.348	**0.745**	cheerful	0.130	0.054	**0.580**
interesting	0.261	0.161	**0.832**	pretty	0.281	0.311	**0.560**
safe	0.440	0.257	0.221	warm	0.371	0.403	**0.755**
favorite	0.375	0.319	**0.707**	full	0.429	0.417	**0.848**
friendly	0.291	0.450	**0.703**	quick	0.158	0.113	0.362
sensitive	0.398	0.306	**0.606**	showy	0.108	0.141	**0.736**
agreeable	0.459	**0.503**	**0.789**	wise	**0.733**	**0.728**	**0.782**
human-like	0.472	0.416	**0.577**	violent	−0.048	−0.156	−0.032
enjoyable	0.150	0.200	**0.759**	good	**0.611**	**0.553**	**0.803**
confident	−0.424	−0.138	−0.203	complex	0.213	0.290	**0.783**
intelligible	**0.780**	**0.569**	0.448	bright	0.155	0.122	**0.741**
approachable	**0.632**	0.286	**0.556**	fast	0.093	−0.003	0.301
considerate	**0.540**	**0.734**	**0.792**	informal	0.235	0.299	**0.563**

Furthermore, ANOVA indicates that there are significant differences in evaluation scores for Q1 and Q2 in every task. Figure 7 shows the graph of means of the evaluations. The multiple comparisons between the scores for both of Q1 and Q2 indicates that the score of Condition C is significantly higher than those of other two conditions in every task. ***, ** and * in Figure 7 represent that there are significant differences with $p < 0.001$, 0.01 and 0.05 respectively as well as Figure 6.

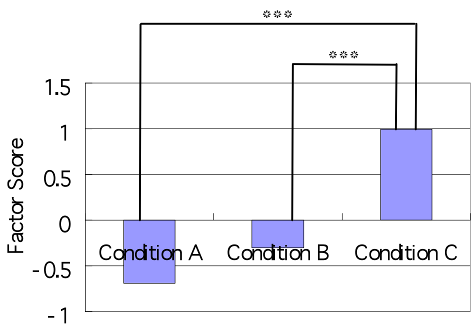
These results of the evaluation of impressions show that the robot which estimates fields of view and behaves with consideration for "difference in perceptions" gives the observer the



(a) Task #1.

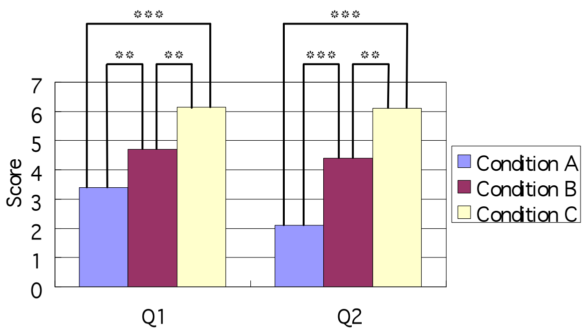


(b) Task #2.

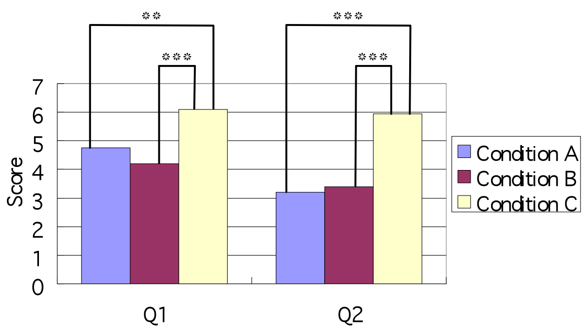


(c) Task #3.

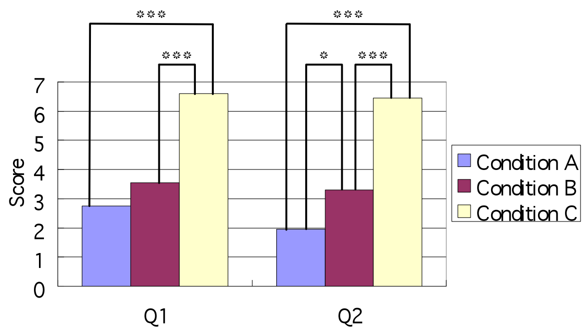
Figure 6: Mean scores for "familiarity."



(a) Task #1 (Third factor).



(b) Task #2 (Third factor).



(c) Task #3 (First factor).

Figure 7: Mean scores for the questions Q1 and Q2.

"familiar" impressions such as "considerate," "agreeable," "approachable" and so on. Furthermore, the results of the two questions Q1 and Q2 show that the participants evaluate the proposed robot's behaviors as appropriate to the situation. The results also show that the robot gives them impressions that it can put itself in the user's position during the interactions.

4. CONCLUSIONS

In this paper, we carry out experimental subjective evaluations of impressions which the robot gives to humans during human-robot interactions. We expect that a "considerate" robot which can guess a user's perception and understand differences between their recognition of situations will be friendlier and more pleasant to users. The experimental results show that such a "considerate" robot can give "familiar" impressions to humans. The contribution of this paper is to verify our expectation statistically by the experimental evaluation from the viewpoint of subjective or Kansei impressions.

The robot which can behave while considering difference in perceptions and putting itself in a user's position has an *imagination* to the user's state and a *consideration* for the user. The experimental results in this paper show that humans can feel such imagination and consideration from the robot's behaviors well and that such abilities contribute toward improving the robot's "familiarity." This fact is expected to be one of fundamental recommendations for designing much friendlier interactions with robots and other intelligent systems.

As a future work, a study which deals with a more complicated "difference in perceptions" should be addressed. In this paper, we focus on the fields of view, which are relatively easy to estimate from the images. Hereafter, we will attempt to study detection of differences in more subjective and Kansei perceptions and to realize a robot which can act more "considerate" behaviors toward humans based on such information. Furthermore, it is probable that the appearance of the robot has a great influence on impressions to users. Therefore it is also required that we carry out subjective impression evaluations using robots of various appearances and analyze the difference in results between them.

ACKNOWLEDGEMENT

This research was partially supported by MEXT Grant-in-Aid for Young Scientists (B) No. 20700182 and the CASIO Science Promotion Foundation.

REFERENCES

1. Sisbot, E. A., Alami, R., Simeon, T., Dautenhahn, K., Walters, M., Woods, S., Koay, K. L., and Nehaniv, C., Navigation in the presence of humans, In the proceedings of HUMANOIDS 2005, Tsukuba, Japan, pp. 181–188, 2005.
2. Dautenhahn, K., Walters, M., Woods, S., Koay, K. L., Nehaniv, C. L., Sisbot, E. A., Alami, R., and Simeon, T., How may I serve you? A robot companion approaching a seated person in a helping context, In the proceedings of HRI 2006, Utah, USA, pp. 172–179, 2006.
3. Pacchierotti, E., Christensen, H. I., and Jensfelt, P., Design of an office-guide robot for social interaction studies, In the proceedings of IROS 2006, Beijing, China, pp. 4965–4970, 2006.
4. Imai, J., and Kaneko, M., Estimation of robot's and user's views and blind regions in occlusion environments, In the proceedings of RO-MAN 2007, Jeju, Korea, pp. 528–533, 2007.
5. Imai, J., and Kaneko, M., Human-robot interaction with view estimation in occlusion environments, In the proceedings of SCIS & ISIS 2008, Nagoya, Japan, pp. 1281–1286, 2008.
6. Doucet, A., de Freitas, N., and Gordon, N., eds., Sequential Monte Carlo Methods in Practice, Springer, New York, 2001.
7. Huynh-Thu, Q., Meguro, M., and Kaneko, M., Skin-color extraction in images with complex background and varying illumination, In the proceedings of WACV'02, Orlando, USA, pp. 280–285, 2002.
8. Lowe, D. G., Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis., Vol. 60, No. 2, pp. 91–110, 2004.
9. Kiryati, N., Eldar, Y., and Bruckstein, A. M., A probabilistic Hough transform, Pattern Recognition, Vol. 24, No. 4, pp. 303–316, 1991.
10. Osgood, C. E., Suci, G. J., and Tannenbaum, P. H., The Measurement of Meaning, University of Illinois Press, Urbana, 1957.