Robots that Can Feel the Mood: Adaptive Interrupts in Conversation Using the Activity of Communications

Akira Imayoshi[1] Hiroshi Yoshikawa[1] Nagisa Munekata[1] Tetsuo Ono[1]

[1] Graduate School of Information Science and Technology Hokkaido University

Abstract: In human communication, "social space" is regarded as a territory formed by a group, and the members of the group may feel upset if they are intruded on by another person without reason. A robot that communicates with humans should observe this social rule. In this paper, we propose a robot system that can recognize the state of a social space by using the smart phones of members of the group. We implemented this system in a robot and did experiments in which we had the robot join in on participants' conversation while changing the timing of its interruption. As a result, the impressions of the participants toward the robot were more favorable when it used the proposed system than when it did not.

1 INTORODUCTION

The "space" surrounding people plays an important role in human communication. It defines the comfortable distance that a person keeps for another person. If two people are too near or too far, communication between them does not go smoothly. In social psychology, the space surrounding an individual is defined as "personal space" [1]. This space is territorial in nature, as stated above. A person tries to be comfortable by keeping his/her space, but if he/she cannot keep it for a long enough time, he/she might feel stressed.

A group also has an inherent space called a "social space", which is any place where people come together and interact with one another [2]. The social space has the effect of strengthening the relationship between the members of the group and rejecting those outside of the group. Social space is the territory of a group, similar to personal space, and the group can comfortably communicate or work by keeping it [3] [4].

A robot that operates in the real world needs to consider personal and social spaces. A robot could easily interfere with the space around humans because it has a body and can move about in the real world. That is, when a robot passes through a social space, it might disturb the group. Moreover, the conversation of the group might be interrupted when a robot talks without understanding the conversation. To avoid this problem, the robot should estimate the sense of distance in a group and the state of the space, and it should select its behavior on the basis of this estimation.

In this paper, we propose a robot system that can recognize the state of a social space by using the smart phones belonging to the members of the group. We implemented this system in a robot and did experiments in which we had the robot join in on participants' conversations while changing the timing of its interruption.

2 RELATED WORKS

As described above, a social space is the territory of a group [2]. The response of a social space to outside changes dynamically depending on the number of members or social relations of the members[3][4].

There have been many studies on recognizing the communication space and social state [5][6][7][8]. These studies have used cameras or sensors embedded in the environment.

Some studies have focused on the interaction between robots and groups. Shiomi et al. propose a method in which a robot determines whether the group is ready listen to it present information to them [9]. A robot gets positional information from floor sensors. It classifies groups using clustering analysis and determines whether each group is orderly or disorderly. Chung et al. propose a method in which a robot recognizes the social space and avoids it [10]. Arai et al. propose a system in which a robot can estimate the timing by which to interrupt a group [11].

Moreover, we can look at the social space from a different viewpoint. When a person acts in some way, the range of perceptions is typically in front of the individual. When people communicate, they tend to share the space between them. Kendon defines this space as O-space [12]. Sharing O-space makes communication more comfortable. Yamaoka et al. propose the model of proximity control for information-presenting robots in consideration of O-space [13].

There are many studies related to robots recognizing social spaces or social relations. However, these systems need to use cameras or sensors embedded in the environment. This approach has problems in that it restricts the range of social space recognition and it has an associated equipment cost. Furthermore, because it is not possible to observe individuals continuously, it is difficult to move a robot in a context-adaptive way. To overcome these problems, we propose a distributed system, composed of smart phones. Each user has a smart phone, and this system locally recognizes a group and its social state.

3 OUTLINE OF THE SYSTEM

3.1 Concept of proposed system

Our system is configured as a multi-agent system. An agent consists of a user and his/her smart phone. An agent gets information from surrounding agents and estimates the social relations between agents. It recognizes a group from these results.

Communication between agents in this system is represented by an effective graph (Figure 1). For example, when the system estimates that the communicative relationship between two users is strong, it strengthens the link between them. On the other hand, when the system estimates that the relationship is weak, it weakens the link. Next, the system estimates groups by using clustering with reference to this effective graph (Figure 2). After that, the system estimates the state of the group by reconfiguring information about the agents in the group (Figure 3). In this study, we simply defined the relationship of communication between agents using only the physical distance between them. Moreover, we defined the state of a group simply from the activity of the conversation.

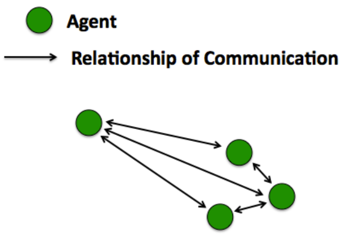


Figure 1: Relationship of communication between agents.

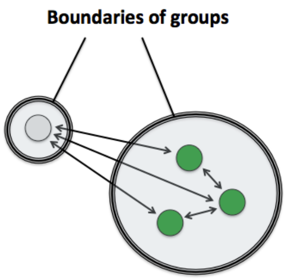


Figure 2: Boundaries of groups.

3.2 Recognition of Social Space

This system recognizes social spaces by using distance information. A user is equipped with a Wi-Fi pocket router and a smart phone. A Wi-Fi pocket router

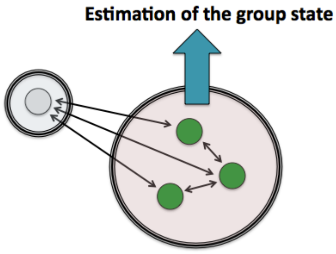


Figure 3: Estimation of the group state

transmits beacon packets that have an SSID corresponding to the agent. A smart phone receives beacon packets and estimates the distance by using RSSI toward a target agent. Communication between agents is performed in P2P mode. If an agent scans a new agent, it accesses a dedicated database server and gets the agent's IP address and a receive port.

We defined the strength of the link between agents as follows.

$$S_{ij}(t) = S_{ij}(t-1) + Input_{ij}(t) - D \qquad (1)$$

$S_{ij}(t)$ is the strength of the link from agent i to agent j at time t. $Input_{ij}(t)$ is the distance to agent j observed by agent i at time t. D is the decay constant.

Each agent builds tables of data by exchanging information with other agents. Table 1 is an example of a table for four agents. The system recognizes groups by using clustering and multidimensional scaling.

	Agent A	Agent B	Agent C	Agent C
Agent A	0	$S_{AB}(t)$	$S_{AC}(t)$	$S_{AD}(t)$
Agent B	$S_{BA}(t)$	0	$S_{BC}(t)$	$S_{BD}(t)$
Agent C	$S_{CA}(t)$	$S_{CB}(t)$	0	$S_{CD}(t)$
Agent D	$S_{DA}(t)$	$S_{DB}(t)$	$S_{DC}(t)$	0

$S_{ij}(t)$: the strength of the link from agent i to agent j at time t.

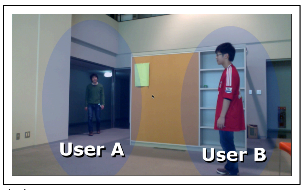
Table 1: Example of a distance table

Next, we describe an example in which the system recognizes the social space shown in Figure 4. In Figure 4-(a), no social space exists because of the large distance between users A and B. In Figure 4-(b), user A approaches user B. Because the distance between them is now small, the system recognizes that they have formed a social space (Figure 4-(c)). In Figure 4-(d), the users have finished their conversation and user A has moved away from user B. In this case, the system recognizes that the social space has disappeared.

3.3 Estimation of Conversation Activity

The activity of conversation is estimated as follows.

(1) Utterance interval detection



(a) Distance between users is large.

(b) User A approaches user B.

(c) System recognizes the social space.

(d) System recognizes that the social space has disappeared.

Figure 4: Example of the proposed system recognizing a social space.

The system detects an utterance interval for input audio data from a user's smart phone (Figure 5).

(2) Value of a voiced interval

The system calculates the amount of utterances of a user for a voiced interval at time t by using the following equation. $Start(i)$ is the start time of the interval, $End(i)$ is the end time of the interval, and α is a constant.

$$H_i(t) = (End(i) - Start(i)) \times \frac{1}{1 + \frac{1}{\alpha}(t - End(i))} \quad (2)$$

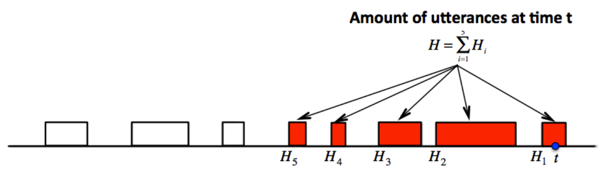


Figure 5: Amount of user utterances at the time t

(3) Calculation for the amount of utterances

The system calculates the amount of utterances by using $H_i(t)$ from step (2). The amount of utterances $U_j(t)$ for user j is calculated as the sum of the last five $H_i(t)$ from the time t.

$$U_j(t) = \sum_{i}^{5} H_i(t) \quad (3)$$

(4) Calculation for determining the activity of conversation

The activity of conversation $A(t)$ is calculated as the mean of $U_j(t)$ of users forming the social space. M is the number of users forming the social space.

$$A(t) = \frac{1}{M} \sum_{j}^{M} U_j(t) \quad (4)$$

4 EXPERIMENT

4.1 Experimental Setup

We did experiments in which a robot intervened in a social space during a conversation. We evaluated the impressions of the participants toward the robot and the influence of the robot on them.

Figure 6 is an outline of the experiment, and Figure 7 shows scenes from the experiments. First, the experimenter had two participants with smart phones sit down on chairs and start a conversation (Figure7-(a)). The robot called out to the participants when the conversation grew lively (Figure7-(b)). To call out to the participants under Condition 1, the robot approached the participants without stopping and began to speak. Under Condition 2, the robot approached the participants, waited, and observed. The robot called out to the participants when there was a chance of making itself heard in the conversation. After that, the robot asked the participants to move the obstacles in front of it. The speech and motions of the robot were the same in both conditions.

We experimented on ten groups of two participants. We analyzed the questionnaires filled out by the participants, videos recorded during the experiments, and the activity of the conversations that resulted from using the system.

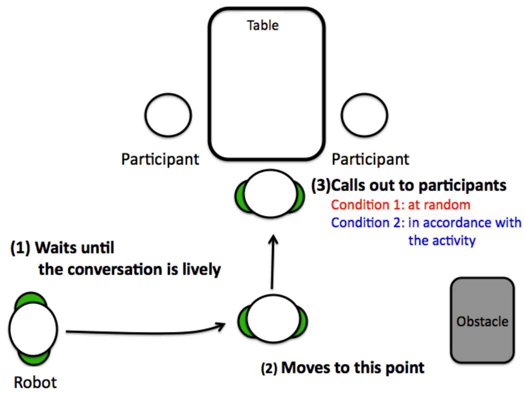


Figure 6: Robot's movements and a setup of the experiment.

4.2 Results

4.2.1 Results of questionnaires

A total of 20 participants evaluated the appropriateness of the timing by which the robot called out to them under Conditions 1 and 2 on a scale of 1 to 7 (Figure 8). For Condition 1, the mean was 3.5, and

(a) Participants talking. (b) Robot interrupting conversation.

Figure 7: Scenes from the experiments

the standard deviation was 1.28. For Condition 2, the mean was 5.1, and the standard deviation was 1.02. A significance test indicated that there was a significant difference between them ($p < .05$).

Moreover, according to the descriptions in the questionnaires, the participants were more favorably inclined toward the robot under Condition 2. In regard to Condition 1 some mentioned: "I was surprised that it called out to suddenly", "The conversation broke off", "I had forgotten what we talked about" and "I didn't expect the robot to talk to me". Regarding Condition 2, some wrote: "I thought that the robot had business with us because it was waiting near us", "The robot waited for our conversation to break off", and "The conversation was not disturbed until we turned our attention to the robot, and our impression of the robot was good".

However, some participants had unfavorable impressions towards the robot under Condition 2. In these cases, the robot had called out to participants, and they felt that it broke off their conversation.

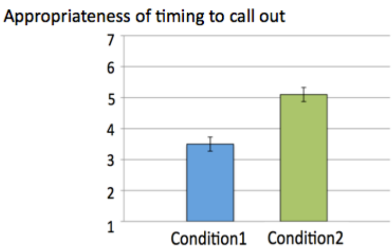


Figure 8: Scores of appropriateness of timing by which the robot called out to participants

4.2.2 Conversation activity and video

Figure 9 and 10 plot the points in time when the robot called out to participants. The activity of the conversation of one group is the vertical axis and time on the horizontal axis. Figure 9 shows the results for Condition 1, and Figure 10 shows those for Condition 2. For Condition 1, Figure 9 shows that the robot called out to participants when the activity of the conversation was high, and the robot caused the activity of the conversation to decrease afterward. For Condition 2, Figure 10 shows that the robot called out to participants when the activity had decreased, and the video shows that participants became aware of the existence of the robot, stopped their conversation, and directed their attention towards it. From these results, how a robot calls out to people could affect their impressions of the robot.

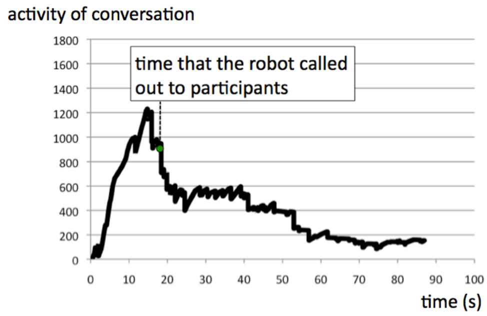


Figure 9: Conversation activity of one group under Condition 1.

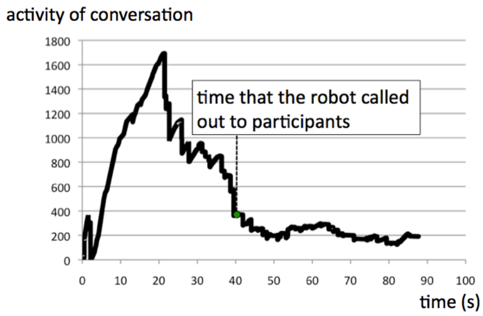


Figure 10: Conversation activity of one group (same group as in Fig. 9) under Condition 2.

5 DISCUSSION

In this study, we developed a method to recognize social space and estimate its state by using a distributed system composed of smart phones. Furthermore, we implemented a system that enables a robot to interrupt a social space in accordance with the activity of the conversation taking place. We conducted an experiment to evaluate the appropriateness of the timing by which the robot called out to participants. In this section, we discuss the problems with this system.

First, the activity of the conversation is insufficient as an index of the timing to call out. In both experiments, good evaluations were obtained when the robot called out at a break in context. However, the evaluations were poor in other cases. For the robot to better "feel" the mood, the system should consider other factors such as what people are talking about.

Second, it is unknown whether similar results would be when the group has more than three people. As the number of a group members increases, the social space widens and the group may have multiple moods.

Third, the performance of recognizing social space was not evaluated. In the future, it will be necessary to determine whether there is any difference between a group that a person(observer) recognizes and one that the system does.

As described above, the system still has certain problems, but it also has a great deal of extensibility. Our system used only the physical distance to recognize a group; however, it can use the other social information such as the direction someone's body is facing or location information using smart phones. The system can also be made to estimate social relationships by using information from social networks, such as Facebook or Twitter, which can not be observed by sensors.

6 CONCLUSION

We proposed a robotic system that uses smart phones to recognize social spaces by using distance information and estimating the activity of a conversation. Subjective experiments on this system showed that participants were more favorable toward a robot when it used this system than when it did not. It seems that it is comfortable for people to accept the behaviors of a robot when the activity level of a conversation is lower, as if the robot could feel the mood.

We conducted experiments on pairs of participants, however, we will do experiments on groups consisting of three or more participants in the future. We will evaluate the proposed method and improve it by incorporating other social information and user action histories.

References

[1] Hall, E. T., "The Hidden Dimension," Anchor Books, 2000, ISBN-10:4622004631.

[2] Shibuya., S., "Hito to Hito tono Kaiteki Kyori - Personal Space toha nanika," 1990, ISBN-10:4140016051 (in Japanese).

[3] Chenyne, J. A. and Efran, M. G., "The effect of spatial and interpersonal variables on the invasion of group controlled territories," Sociometry, 1972, 35, 477-489.

[4] Lindskold, S., Albert, K. P., Bear, R., and Moore, W.C., "Territorial boundaries of interacting groups and passive audiences," 1976, 39, 71-76.

[5] Aggarwal, J. K. and Ryoo, M. S., "Human Activity Analysis: A Review," ACM Computing Surveys, Vol.43, No.3, Article 16 (2011).

[6] Pentland, A., "Social Signal Processing," IEEE Signal Processing Magazine, Vol.24, No.4 pp.108-111 (2007).

[7] Turaga, P. T., Chellappa, R., Subrahmanian, V. S. and Udrea, O., "Machine Recognition of Human Activities: A Survey, " IEEE trans. Circuits and Systems for Video Technology, Vol. 18, No. 11, pp. 1473-1488 (2008).

[8] Vinciarelli, A., Pantic, M. and Bourlard, H., "Social Signal Processing: Survey of an Emergent Domain, " Image and Vision Computing, Vol. 27, No. 12, pp.1743-1759 (2009).

[9] Shiomi, M., Nohara, K., Kanda, T., Ishiguro, H. and Hagita, N., "Estimating Group States For Interactive Humanoid Robots," Humanoid Robots, 2007 7th IEEE-RAS International Conference, p.318-323.

[10] Chung, S. and Huang, H., "Incremental learning of human social behaviors with feature-based spatial effects," Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference, p.2417-2422.

[11] Arai, T., Takahashi, K. and Kaneko, M., "Interruption Strategy Delivery Robots Aware of Receiver's Condition," The Institute of Image Information and Television Engineers, ITE Technical Report Vol. 36, No. 8, ME2012-36 (Feb. 2012).

[12] Kendon, A., "Conducting Interaction-Patterns of Behavior in Focused Encounters," Cambridge University Press, 1990.

[13] Yamaoka, F., Kanda, T., Ishiguro, H and Hagita, N., "A Model of Proximity Control for Information-presenting Robots, " Journal of the Robotics Society of Japan, 27(2), 230-238, 2009.