How Many Social Robots Can One Operator Control?

Kuanhao Zheng[1, 2]   Dylan F. Glas[1]   Takayuki Kanda[1]   Hiroshi Ishiguro[1, 2]   Norihiro Hagita[1]

1. Intelligent Robotics and Communication Laboratories	2. Intelligent Robotics Laboratory
ATR	Osaka University
2-2-2 Hikaridai, Keihanna Science City	1-3 Machikaneyama Toyonaka
Kyoto, 619-0288 Japan	Osaka, 560-8531 Japan

{tei, dylan, kanda, ishiguro, hagita}@atr.jp

ABSTRACT

This study explores the nature of the multi-robot control problem for social robots. It begins by modeling the overall structure of a human-robot team for social interactions, and implements it for specific application to dialog-based interactions. Operator activity during control of a social robot is studied. Customer satisfaction is proposed as an important metric for evaluating the performance of a human-robot team for social interactions with customers. Based on the modeling, fan-out of a social robot team can be calculated, and the performance of the team is estimated by simulation. A field trial was conducted in a shopping mall to demonstrate a successful deployment of social robots for a real-world application with ensured performance prior to installation using our modeling and simulation approach.

Categories and Subject Descriptors

H.5.2 **[Information Interfaces and Presentation]**: User Interfaces - *Interaction styles*; I.2.9 **[Artificial Intelligence]**: Robotics

General Terms

Design, Human Factors

Keywords

Human-robot interaction, modeling, simulation, social robots

1. INTRODUCTION

Research and development of social robots in human-robot interaction has been making rapid progress in recent years [1, 6, 9, 17, 20, 21, 27]. Social robots have been placed in museums [2, 22, 24, 26], reception areas [16], shopping malls [14, 15], and transit stations [12], providing valuable resources for social robot study. However, current technology is not mature enough to achieve a fully automatic robot in a real-world environment. A speech recognition system that performed with 92.5% accuracy in 75dBA noise [13] achieved only 21.3% accuracy in a real environment [23]. One practical approach to solving this problem is teleoperation, in which human perception and intelligence can fulfill the technology gap in automation.

Teleoperation of multiple robots can be achieved by switching an operator's attention among robots with partial autonomy, which improves system efficiency by enabling multiple robots to act simultaneously. However, the coordination between autonomy and operation is often difficult in real applications of social robots. How many robots can be deployed? How much effort should be expended on automation and how well must the operator perform? Previous studies didn't address such issues, and deployment of social robots was often best-effort-based or ad-hoc, causing unnecessary time and effort for preparation or even failure of the robot team when performance couldn't be estimated prior to deployment.

Teleoperation of multiple robots is an active field of research for search-and-rescue and navigation, from which we can learn useful theories such as adjustable autonomy [4, 10], mixed initiative control [11], fan-out [18, 19] and metrics for robot team performance [3, 5]. However, apart from those applications, social robot study has unique features, such as low tolerance of customers to robot failure [14] and the special structure of social interactions [7], which encourage us to explore the nature of teleoperation for social robots with new considerations.

In this paper, we start by building a top-level model for the scope of our study, and implement it for dialog-based interactions as a concrete application of theory. Simulation is introduced as a strong tool for exploring various factors in our theory. Finally, we apply and test our findings with a field trial in a real environment.

2. OVERVIEW

Figure 1 illustrates the basic workflow of a social human-robot team (HRT). We study an HRT consisting of a single operator and a certain number of semi-autonomous robots to conduct social interactions with customers. Here, we use "customer" to mean a person who engages in social interaction with a robot, as opposed to the operator of the robot.

The operator's job is to control robots to interact with customers through a teleoperation interface for the situations which cannot be handled automatically by the robot, or when a high risk of error exists in the autonomous system. To perform an operation, the operator should acquire situation awareness about the interaction happening between the customer and robot, and give proper inputs through the interface to control robot behaviors.

As the operator is partially needed by each robot, teleoperation of multiple robots can be achieved by switching the operator among multiple robots. Switching efficiency is critical for the task accomplishment of an HRT, as it determines whether the operator can be assigned in time to the robots needing operation.
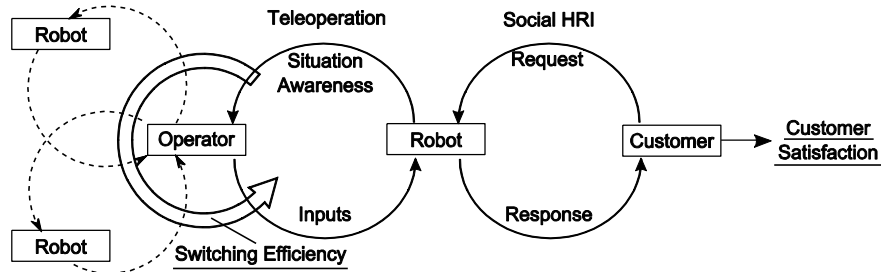
Figure 1. Basic workflow of a social human-robot team

The existence of customers is a unique feature of social HRI. Interactions progress by the exchange of requests and responses between customer and robot; typically, the customer makes a request and the robot's task is to respond to the customer in a socially acceptable way. In this study, we assume such interactions are dialog-based, with a customer asking a question and the robot answering, and each dialog contains only one pair of question and answer.

We define **Customer Satisfaction** as a quantitative evaluation of a customer's emotion after interacting with a robot. For applications of HRT providing a service by interaction with customers, customer satisfaction can be applied as an important metric for measuring the performance of robots.
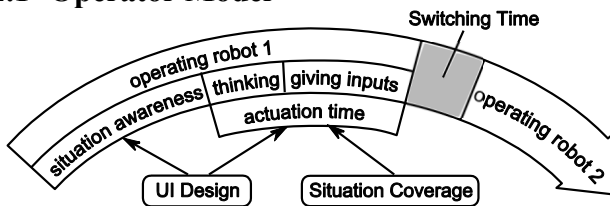
2.1 Operator Model

Figure 2. Operator model

The operator model is explained in Figure 2, which illustrates the activities of the operator when operating multiple robots in sequence, and the key terms that are related to operation time.

As in [3], operation time for a robot can be separated into time for (a) attaining situation awareness, (b) thinking of proper inputs and (c) giving inputs via the interface. We combine (b) and (c) as actuation time, which enables us to measure it in experiments. Between the operations of each robot, there may be an interval of switching time for deciding which robot to switch to.

The UI design affects operation time by affecting time for both situation awareness and actuation. A well-designed UI should help the operator to quickly be aware of the situation about which the customer has asked, and enable quick actuation for responding by providing adequate and efficient inputs.

Situation Coverage (SC) is a metric for task difficulty defined as the percentage of prepared responses among all interactions with customers [8], which affects actuation time. A covered situation means the robot's response behavior and the corresponding UI input to trigger that response are prepared, so that actuation can be made quickly; an uncovered situation means the robot does not have a prepared response behavior, which results in slow actuation while the operator makes an improvised response. Even for a well-prepared system, uncovered situations may occur since customer questions can be difficult to predict.
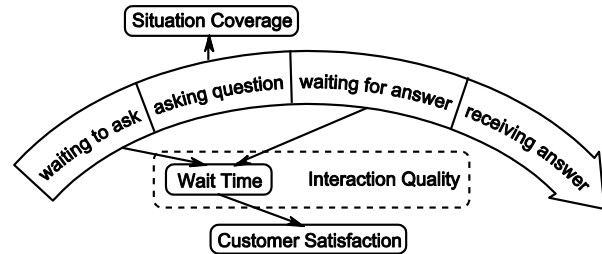
2.2 Customer Model

Figure 3. Customer model

Figure 3 depicts a typical flow of customer activities during a dialog, which is composed of four steps: (a) waiting to ask, (b) asking a question, (c) waiting for an answer and (d) receiving the answer.

Customer satisfaction is affected by various factors that we refer to as interaction quality, such as sociability and friendliness of a robot, and accuracy and promptness of its response. Among these factors, we focus on the effect of customer wait time during an interaction. Wait time comes from a delay of operation, which is a unique feature in the study of teleoperation of social robots. In this study, we assume the accuracy of responses can be ensured by the operator, and examine how customer satisfaction changes for different wait times.

There are two types of wait time in an interaction: waiting to ask a question and waiting for an answer. Research shows that customers get frustrated while waiting [25], and we believe that the effects of the two wait times are different, which will be explored in detail in this study.

Situation coverage connects the customer and operator models, as it is determined by the content of customer questions, and it affects operation time. This means customer activities may affect an operator's performance because the operator's task is to respond to customers when controlling a social HRT.

3. MODEL IMPLEMENTATION

We build a concrete model for HRTs engaged in dialog-based interactions in which the operator works as a speech recognizer and provides expert knowledge. With such models, an efficient operator assignment algorithm is achieved based on the anticipation of interaction structure and operation time.

3.1 Customer-Robot Interaction Model

A previous study has provided a way of modeling conversational interactions based on priorities of interaction phases [7]. By implementing this on the customer model presented in section 2.2, we define phases of interaction as shown in Figure 4.
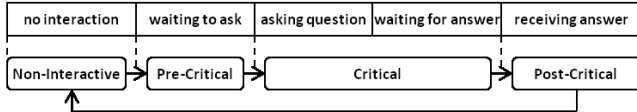
Figure 4. Customer-robot interaction phases

A conversation between customer and robot is modeled by four phases which vary in time-criticality:

- **Non-Interactive Phase:** This is when no interaction occurs and the robot is waiting for a customer's arrival.
- **Pre-Critical Phase**: This is when a customer arrives and is waiting to ask a question before the robot finishes talking, such as a greeting or self-introduction by the robot.
- **Critical Phase:** This phase includes a customer asking a question and waiting for an answer. The operator conducts speech recognition in this phase, and may provide expert knowledge for uncovered situations.
- **Post-Critical Phase:** This phase is when the operation is complete and the robot answers by executing designated behaviors.

The critical phase is important for the operator's task assignment because in this model, it is the only time during an interaction when the operator is needed, and latency of operation may cause failure of an interaction because of the customer's frustration. All the other phases can be handled automatically by robot.

3.2 Model of Customer Satisfaction

We hypothetically model the drop in customer satisfaction as a linear function of wait times before and after asking a question when assuming a consistent quality of answers. As in equation (1), customer satisfaction (S) has an initial value V, which drops in speeds α and β during wait times before (t_{before}) and after (t_{after}) asking a question.

$$S = V - \alpha t_{before} - \beta t_{after} \qquad (1)$$

α and β denote different drop speeds of satisfaction in two conditions of waiting, which are also affected by the context of interaction. For example, while waiting for the same amount of time, a customer will be less frustrated when waiting for an answer that is inherently hard to answer than for an easy-to-answer one. We believe $\alpha < \beta$ in general because the time waiting for an answer is more critical than the time waiting before a question, causing more anxiety to the customer. This hypothesis will be verified by data collection from customers.

Two mechanisms can be applied from previous studies to mitigate customer frustration during wait time, which we refer to as *Wait-Time Management*:

1) Proactive Timing Control (PTC)
This technique is introduced in [8] for controlling the length of the pre-critical phase. When the operator is busy with tasks of other robots and cannot be assigned to a robot entering the critical phase, the robot can perform PTC to delay the entrance to the critical phase by talking about prepared content for a planned amount of time. As the robot keeps talking within its "turn" in the conversation, the extra behaviors may provide a natural impression to the customer and cause less frustration.

2) Conversational Filler
This is introduced in [25] to mitigate a customer's frustration while waiting. It can be used in the critical phase when a customer is waiting for an answer. During such time, the robot can speak phrases such as "well...", "let me think...", or "uh..." to give the impression that it is actively thinking of an answer. Experiments [25] show that conversational fillers successfully moderate customers' negative impression of a long waiting time.

3.3 Model of Operation Time

Operation time is critical for a social HRT because a customer should continue waiting while an operation is completed. Starting with the model in section 2.1, we define operation time ($t_{operation}$) as the sum of listening time (t_{listen}) and actuation time ($t_{actuation}$), as in equation (2). The time for attaining situation awareness is represented as listening time because audio is the major source of information in dialog-based interactions.

$$t_{operation} = t_{listen} + t_{actuation} \qquad (2)$$

Reducing Listening Time
Listening time is a function of the utterance time required for the customer to ask a question. We can buffer the audio during dialog and reduce the operator's listening time by playing back with acceptably faster speed. Let K denote playback speed, t_{ask} denote customer asking time, and t_{absent} denote the absent time between when the customer starts asking and the operator starts listening, then t_{listen} can be calculated by equation (3). The maximum is taken because listening can't end before a customer finishes asking a question.

$$t_{listen} = \max\{t_{ask} - t_{absent}, t_{ask}/K\} \qquad (3)$$

By reducing listening time, total operation time can be reduced, which results in higher customer satisfaction by making a customer wait less during an operation.

Situation Coverage and Actuation Time
As introduced in section 2.1, situation coverage affects an operator's actuation time by influencing the operation interface. For dialog-based interactions, controls such as buttons can be used to trigger prepared utterances for covered situations; otherwise, text-to-speech can be used for generating utterances for uncovered situations, but this may take longer actuation time as the operator must type the entire phrase for an answer. The differences in actuation time for different input types are studied in section 3.5.

3.4 Switching Algorithm

A challenge in teleoperation of multiple robots is timely allocation of the operator to robots requiring operation, which is referred to as switching efficiency. We propose a switching algorithm based on our interaction models, using estimates of operation time for multiple input mechanisms.

3.4.1 Basic Mechanism

To ensure that the operator is always assigned to the robot whose customer has been waiting the longest time in the critical phase, two FIFO queues are maintained according to phase start time, as shown in Figure 5. The operator is always assigned to the robot at the head of the Critical Queue, which is the first robot to start the critical phase. Other robots perform PTC until the operator is assigned to them.
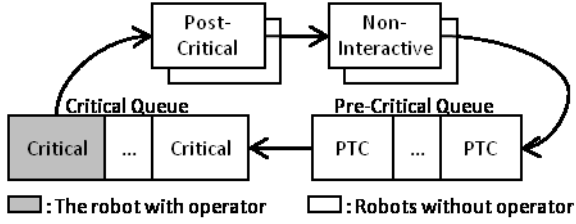
Figure 5. Queuing of robots by switching system

With fast-playback, the critical phase can start ahead $(1 - 1/K)$ of asking time before the operator is assigned because listening time is $1/K$ of asking time, which results in the following decision logic, where $\sum \tilde{t}_{operation}$ denotes the sum of estimated operation time in the Critical Queue:

> IF ($\sum \tilde{t}_{operation} \leq (1 - 1/K) \cdot \tilde{t}_{ask}$)
>> Proceed to critical phase.
> ELSE
>> Continue PTC.

3.4.2 Estimation of Operation Time

When multiple input types exist and it is impossible to tell which type will be used before operation, actuation time can be estimated using a probabilistic approach. Suppose there are n input types, each with a probability p_i being used and a Gaussian distribution $N(\mu_i, \sigma_i^2)$ of actuation time. A penalty function can be defined by equation (4), where x, y denote estimated and actual actuation times, and α, β are penalties of wait time as in customer satisfaction equation (1).

$$Q(x,y) = \begin{cases} \alpha(x - y), \text{when } x > y \\ \beta(y - x), \text{otherwise} \end{cases} \quad (4)$$

Let $E_i[Q(t)]$ denote the expected penalty for input type i when estimated actuation time is t; then it is as in equation (5), where $f(x; \mu_i, \sigma_i^2)$ is the probability density function. The expected penalty for estimation t is the expectation from all possible input types, as in equation (6). Estimated actuation time ($\tilde{t}_{actuation}$) is taken with minimum expected penalty, as in equation (7).

$$E_i[Q(t)] = \int_0^\infty f(x; \mu_i, \sigma_i^2) Q(t, x) dx \quad (5)$$

$$E[Q(t)] = \sum_{i=1}^n p_i E_i[Q(t)] \quad (6)$$

$$\tilde{t}_{actuation} = \arg \min_{0 < t < \infty} E[Q(t)] \quad (7)$$

Best estimation can be achieved by comparing expected penalties for all possible estimations. For example, suppose the operation of three input types have the same probability $(1/3)$, and actuation time of each type are $N(1.9, 0.6^2)$, $N(3.1, 1.9^2)$, $N(32.9, 11.9^2)$ seconds. Figure 6 shows the expected penalty for different estimations, where 22 seconds is the best estimation.
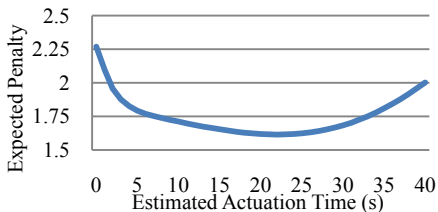

Figure 6. Estimation of actuation time and expected penalty

3.5 Data Collections

Data collections were conducted to get human data for our models of the customer and operator. To investigate the impact of complexity in dialog-based interactions, two scenarios were used.

- **Guide Scenario:** Robots work at a shopping mall to provide route guidance service. Customers ask questions about where some shops are, and robots answer accordingly.
- **Seller Scenario:** Robots work as sellers at a PC shop. Customers ask various questions related to PCs or peripherals, and robots give accurate answers.

The first scenario represents a context when customers are in a hurry and interactions are short. The second one is about a relatively complex scenario, when customers are not in such a hurry but need detailed information.

Portrait of Participants
Undergraduate Japanese students were recruited for data collections regardless of whether they had any background in robotics. We did not allow the same participants to take part in both data collections for operator and customer, because knowing how robots are operated may affect a participant's evaluation when acting as a customer. Basic computer skills for daily life were required for participants acting as operators.

3.5.1 Customer Data Collection

Fifteen people participated (8 female, 7 male, mean 22 years old) as customers, asking 16 questions for each of two scenarios. The robot acted with PTC among 0, 15, 30, 45 seconds and delay with conversational fillers among 0, 5, 10, 15 seconds. Satisfaction was scored with integers from -5 to 5, where -5 and 5 indicate maximum negative and positive satisfaction. Each scenario was repeated twice to counter-balance the ordering effect.

By linear regression analysis using least squares, parameters of equation (1) were calculated as in Table 1 (decision coefficient R^2 are 0.970 and 0.967 for each scenario, which indicate very good fitting). Asking and answering times were measured, where μ and σ are mean and standard deviation.

Table 1. Parameters from customer data collection

Scenarios	V	α	β	Customer Asking Time (s)		Robot Answering Time (s)	
				μ	σ	μ	σ
Guide	3.65	0.07	0.18	4.1	1.4	5.1	0.6
Seller	3.68	0.04	0.14	5.8	1.8	10.6	1.7

It was verified that $\alpha < \beta$ for both scenarios, meaning people are generally more patient when waiting before than after asking. With larger V, smaller α and β, customers tend to be more patient in a seller scenario, meaning people's patience is different depending on the complexity of conversation.

3.5.2 Operator Data Collection

Sixteen people participated (7 female, 9 male, mean 21 years old) in the two scenarios. To explore the effect of UI on actuation time, operation was measured using three input types: binary choice, list choice and typing. For binary choice, two choices were shown; for list choice, 20 choices were shown; and for typing, the operator directly entered the answer into a text field.

Table 2 shows the result. Actuation time increased as the complexity of operation increased, with typing time obviously longer than the other two, and selection from list choices took longer than binary choice. The operation of the seller scenario took longer than the guide scenario for each interface, revealing the effect of conversation context on actuation time.

Table 2. Measured actuation time for different input types

Input Types		Binary		List		Typing	
		guide	seller	guide	seller	guide	seller
Actuation	μ	1.9	2.2	3.1	5.5	32.9	45.0
Time (s)	σ	0.6	0.9	1.9	4.8	11.9	18.5

4. SIMULATION

This modeling provides the potential for studying social human-robot interactions in great detail, which we can achieve by simulation. We built a simulation tool for estimating the performance of an HRT, and conducted an experiment to validate its result. Then, we explored the effects of different strategies and techniques in teleoperation of social HRTs using simulations.
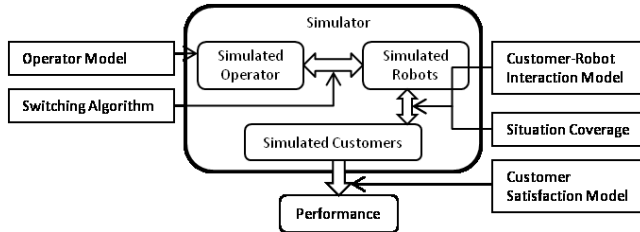
4.1 Simulation Tool

Figure 7. Outline of simulation tool

We built a simulation tool to simulate the process of interactions without actual participation of humans or robots. As illustrated in Figure 7, it is a computer program simulating activities of an operator, robots, and customers, which runs based on time by refreshing the status of simulated entities, triggering events representing dynamics of interaction phases and operator activities with specified timings.

Customer-robot interactions are simulated by updating interaction phases with durations of customer and robot utterances in each phase, which can be measured from real interactions. Operation time of each simulated robot is calculated by the operator model, based on asking time from the simulated interaction and specified actuation time. The switching algorithm decides the mapping from the simulated operator to robots just as in a real application.

As the output of the simulation, customer satisfaction for each interaction is calculated by counting wait times of the simulated customers during PTC and the critical phase. Performance is calculated as the sum of satisfactions from all interactions divided by total simulation time, as in equation (8).

$$Performance = \frac{\sum Satisfaction}{Time} \quad (8)$$

4.2 Validation Experiment

The experiment was conducted to determine whether simulation can provide a reliable result in comparison with human operators. Fifteen people participated (6 female, 9 male, mean 20 years old).

4.2.1 Procedure

The validation was processed by comparing the performance of robot teams (a) operated by participants and (b) from simulation, for each team size from 1 to 8. If two conditions get similar results for each robot number, it can be verified that simulation can provide trustworthy estimation.

For condition (a), real robots and customers were not set up, but recorded audio of questions was used to simulate customers. We didn't recruit people to be customers because audio of their questions can give the operator a similar experience of teleoperation, and by using the wait time resulting from operation, performance can be calculated the same as with customers involved. For condition (b), data measured from interactions as in Tables 1 and 2 were used as inputs, which enables the simulator to generate timings based on real-world parameters in its execution.

A guide scenario as described in section 3.5 was set for both conditions, and frequent customer arrival was simulated by setting the arrival interval as $N(5,2^2)$ seconds. Situation coverage was set to 90%, with list choice and typing available for covered and uncovered situations.

4.2.2 Results

Figure 8 shows the comparison of performance[1] from simulation and mean values by participants, where standard deviations of participant data are also depicted. Although slight differences in some data points exist due to variation of performance by participants, the changes of performance show the same trend, and both results indicate the fan-out being 3 by forming performance plateaus of similar shapes. Thus, we can conclude that simulation provides reasonable estimation regarding actual performance when using data measured from real interactions.
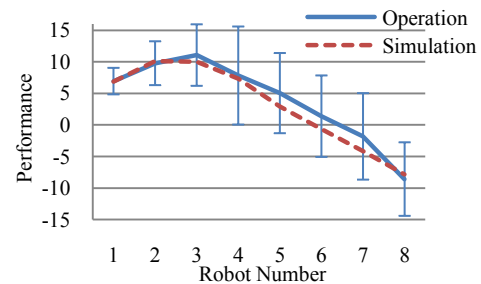

Figure 8. Comparison of performance between human operators and simulation

4.3 Explorations with Simulation

Once it is proved that simulation can provide valid estimations, the effects of various strategies and techniques can be explored using simulations without actual dispatch of human-robot teams.

4.3.1 Effect of Estimation and Fast-Playback

To explore the effect of operation time estimation and fast-playback for improving performance of an HRT, simulations were run under four conditions. As the baseline condition, neither estimation nor fast-playback was used, so that PTC will last until an operator is assigned, and listening was at normal speed. For

---

[1] The time unit for measuring performance is "Minutes" throughout this paper.

conditions with only estimation or fast-playback, the effects of using those single techniques were explored. The condition of using both techniques is the same as our switching algorithm. Situation coverage and customer arrival were set to be the same as in subsection 4.2.

As seen in Figure 9, conditions using neither technique and using only fast-playback get the same result, and conditions with only estimation and with both techniques get gradually improved performance. This shows that expectation of operator activities and reducing operation time are both helpful for improving performance. It also shows a case where a technique (fast-playback) for reducing operation time is not effective until good estimation of operation is applied.
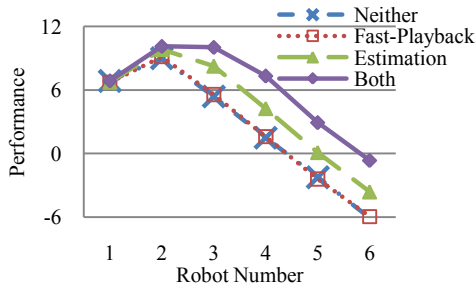

Figure 9. Comparison of performance by four conditions

4.3.2 Effect of PTC

We simulated the effect of using and not using PTC. When not using PTC, customers have to wait after asking even if the operator is busy with other tasks. With PTC, it is the same as our switching algorithm. Situation coverage and customer arrival were set to be the same as in subsection 4.2.

As shown in Figure 10, the two conditions get similar results up to two robots, and the performance without PTC severely drops for larger team sizes. Fan-out is increased by 1 with PTC. More importantly, the results indicate that PTC can effectively mitigate the risk when customers simultaneously arrive at many robots.
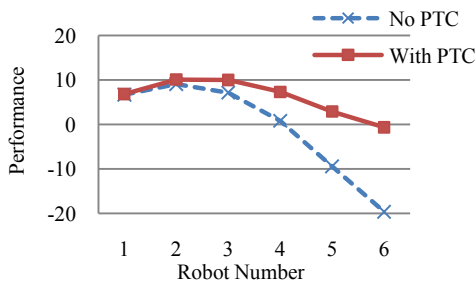

Figure 10. Comparison of using and not using PTC

5. FIELD TRIAL

A field trial was then conducted in a shopping mall over four days. The goal was to use the modeling strategy to prevent ad-hoc settings for preparation and ensure adequate performance of the human-robot team in a deployment for a real-world application.

5.1 Setup and Procedure

We were asked to provide service with robots at an event for the anniversary of a shopping mall, including the tasks of (a) providing route guidance service, (b) playing with children and (c) replying to customers' questions. ROBOVIE-II communication robots were used to interact with customers.

To prepare the system for the tasks, our procedure for the field trial is described in Figure 11. It consists of three steps, and simulation was used throughout all steps.
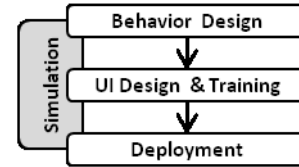

Figure 11. Procedure of field trial

At the beginning step, robot behaviors for answering customers' questions were designed and implemented. Although designing more behaviors is helpful for improving performance, the time and effort to design behaviors are precious and should be preserved; also, we can't spend an infinite amount of time for preparation in reality. Thus, a decision to set a target amount was necessary, which was achieved by simulation using acquired information.

The user interface was designed for implemented behaviors, and an experienced operator was assigned during the whole procedure. Operation time is critical for a social HRT, which affects the wait time of a customer; but how quickly should an operator respond? Simulation was used to estimate performance with measured operation time, and improvements of UI and training were made until we could expect an acceptable performance by the operator.

Deployment was conducted when good performance could be expected after the previous steps. We deployed the proper number of robots to ensure good performance, and observed the activities of the robot team interacting with customers.

5.2 Results

5.2.1 Effects of Modeling

Our modeling of customer-robot interactions and operation contributed to making important decisions in the deployment procedure when simulation was applied based on the modeling. Advantages that make our deployment different from previous approaches are explored.

Effect for Target Setting

At the beginning step, the target of team size and necessary situation coverage (SC) were calculated using the model of interactions we studied for the guide scenario.

Table 3 shows the lower boundaries of SC for controlling different numbers of robots in simulation using data collected for the guide scenario. The result indicates at most three robots can be controlled when SC is over 91%, and 41% is needed to dispatch at least one robot.

Table 3. Minimum required SC for different fan-out

Fan-out	1	2	3
Minimum SC	41%	74%	91%

Since a previous field trial in a shopping mall [14] achieved over 90% SC, we set the target to deploy three robots by implementing behaviors to achieve 91% SC. To get a high level of SC, we made (a) 99 behaviors for route guidance, (b) 6 behaviors for playing with children, and (c) 17 behaviors for providing information.

We needed to measure whether behaviors could provide the expected SC from real interactions prior to deployment because insufficient behaviors may cause failure in the real deployment due to low SC. Good performance should be ensured in this procedure because customers will be involved. One robot was used for measurement because simulation tells us only 41% SC is needed for it to perform well, which we believed could be achieved using the implemented behaviors. SC was measured as 95.7%, higher than our expectation. Hence, we interpreted this to mean that the preparation of behaviors was sufficient.

The target for preparation and needed effort to reach the target was successfully estimated using simulation; otherwise, we would not be able to know how many robots to deploy and how many behaviors to implement in the preparation step.

Effect for Performance Estimation

Before robots were dispatched for measuring SC, a UI was implemented for the new behaviors, and the operator was trained until an acceptable performance could be expected.

When the UI was implemented, actuation time for list and map inputs were measured, and simulation of the robot team with different sizes was run with measured actuation time, as shown in Figure 12 labeled as "before". The simulation shows fan-out is one with performance barely above zero, meaning the HRT would perform very badly if deployed with the current condition.
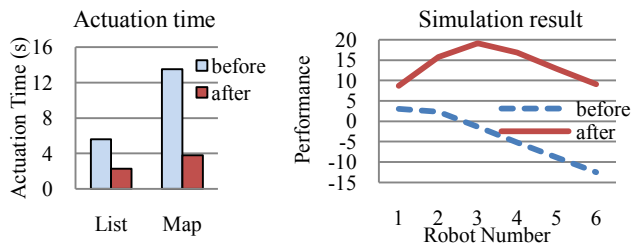

Figure 12. Actuation time and simulation result before and after improvement

This low performance resulted from long actuation time, especially for the map interface. From the operator's feedback, we found that operation with the map was difficult because opaque buttons inhibited reading of the map. To solve this, we made the buttons semi-transparent, as in Figure 13, so that the operator could easily understand the locations represented by the buttons.
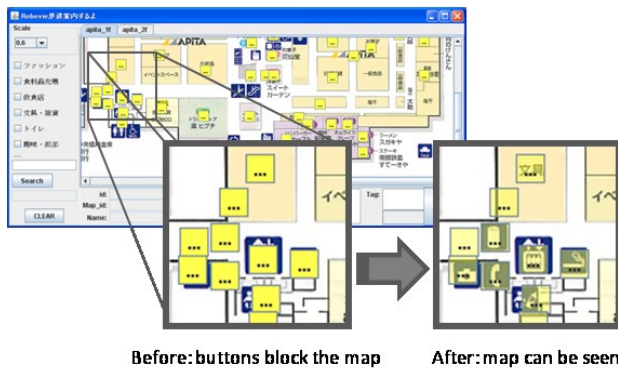

Figure 13. Map UI before and after improvement

Actuation time was measured after training with the improved map UI. The simulation was run using the updated actuation time. As shown with the label "after" in Figure 12, fan-out increased to

three, indicating the end of preparation by accomplishing our initial target.

By performance estimation, we were able to determine acceptable actuation time, and a potential failure of deployment by deficiency of UI was prevented. This is different from previous approaches, where problems were found and solved through repeated trials in a real environment. In our approach, potential problems can be found and solved before trials, which is very important for ensuring quality of service provided by social robots.

5.2.2 Deployment Results

One robot was deployed in the first two days because few customers were expected, and three robots were deployed in the last two days. Figure 14 shows three robots simultaneously interacting with children customers.

Figure 14. Three robots interacting with children

Situation coverage remained over 99% among a total of 285 interactions conducted over the four days, indicating that prepared behaviors were sufficient to respond to most customers' requests. As evaluation of performance, we measured customer wait times before and after questions, as shown in Table 4. When three robots were deployed in the last two days, mean wait times before and after questions increased by no more than 2 seconds, while many more interactions were conducted as the payoff from using multiple robots.

Table 4. Customer wait time and number of interactions

Days		1	2	3	4
Mean Customer Wait Time (s)	Before[2]	0.0	0.0	1.2	2.0
	After	3.6	3.9	5.3	5.0
Number of Interactions		34	26	117	108

This modeling approach led to a successful field trial. By simulations based on the modeling, time and effort for preparation were saved by setting the proper target at the beginning, and the quality of service was ensured before the robots met customers, which is a merit for applications in which customers have low tolerance for bad service.

6. DISCUSSION

6.1 Summary

We have achieved practical modeling of social HRT for dialog-based interactions. Operation time is modeled by listening and actuation. Customer-robot interaction is modeled by phases. Customer satisfaction is based on wait time during interaction, which is used for measuring performance of social HRT. Situation coverage is revealed as a connection between the customer and operator model, which is the result of customer activities and affects operation time. The connection between models and merits of modeling were validated through a field trial.

---

[2] The wait time before questions doesn't include the time for a robot to do necessary self-introduction before each interaction.

With the model, we can answer the question of "how many social robots can one operator control" by computing fan-out based on performance estimation. By simulation using real-world data as inputs, the quality of service provided by a robot team could be estimated before the actual installation. We believe that this study provides a powerful method of designing a teleoperation system for controlling multiple social robots.

6.2 Limitations

One limitation of this paper is that the parameters for our models were set for a specific context, where robots worked in a defined area and interactions were restricted to a few topics. However, we believe the models can be applied to a large range of problems by adjusting parameters from real-world data for specific contexts.

Also, our model only covered interactions with single round question and answer. For interactions with multiple questions, the customer-robot interaction model can be extended by adding looping of dialog and applying different words to robots for first-round and consecutive dialogs.

Another limitation is that our study did not model random errors of automation. In our model of interactions, we predict the moment when error could happen as the critical phase, but in fact, random errors may happen any time, both for conversation (e.g. sudden departure of customer) or navigation (e.g. robots hit obstacle). In the future, as speech recognition has the potential to make parts of interactions fully autonomous, the possibility of errors caused by this should also be considered.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Breazeal, C., et al. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. IROS2005, pp. 383-388, 2005.

[2] Burgard, W. et al. The interactive museum tour-guide robot. Proc. of National Conference on Artificial Intelligence, pp. 11-18, 1998.

[3] Crandall, J. W. and Cummings, M. L. Developing performance metrics for supervisory control of multiple robots. HRI2007, pp.33-40, 2007.

[4] Crandall, J. W. and Goodrich, M. A. Characterizing efficiency of human robot interaction: a case study of shared-control teleoperation. IROS2002, pp. 1290-1295, 2002.

[5] Crandall, J. W., Goodrich, M. A., Olsen, D. R. and Nielsen, C. W. Validating human-robot systems in multi-tasking environments. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 35(4): 438–449, 2005.

[6] Dautenhahn, K., et al. How may I serve you?: a robot companion approaching a seated person in a helping context. HRI2006, pp.172–179, 2006.

[7] Glas, D. F., Kanda, T., Ishiguro, H. and Hagita, N. Field trial for simultaneous teleoperation of mobile social robots. HRI2009, pp.149-156, 2009.

[8] Glas, D. F., Kanda, T., Ishiguro, H. and Hagita, N. Simultaneous Teleoperation of Multiple Social Robots. HRI2008, pp.311-318, 2008.

[9] Gockley, R., Forlizzi, J., Simmons, R. Natural person following behavior for social robots. HRI2007, pp. 17-24, 2007.

[10] Goodrich, M. A., et al. Managing autonomy in robot teams: observations from four experiments, HRI2007, pp. 25-32, 2007.

[11] Hardin, B., Goodrich, M. A. On using mixed-initiative control: a perspective for managing large-scale robotic teams, HRI2009.

[12] Hayashi, K. et al. Humanoid robots as a passive-social medium - a field experiment at a train station-. HRI2007, pp. 137-144, 2007.

[13] Ishi, C. T. et al. Robust speech recognition system for communication robots in real environments. IEEE Int. Conf. on Humanoid Robots, pp. 340-345, 2006.

[14] Kanda, T., et al. An affective guide robot in a shopping mall. HRI2009, pp. 173-180, 2009.

[15] Kanda, T. et al. Who will be the customer?: A social robot that anticipates people's behavior from their trajectories. UbiComp2008, 2008.

[16] Michalowski, M.P., Sabanovic, S. and Simmons, R. A spatial model of engagement for a social robot. Proceedings of AMC 2006.

[17] Mutlu, B., Hodgins, J. K. and Forlizzi, J. A storytelling robot: modeling and evaluation of human-like gaze behavior. IEEE Int. Conf. Humanoid Robots, pp. 518-523, 2006.

[18] Olsen, D. R. and Goodrich, M. A. Metrics for evaluating human-robot interactions. NIST's Performance Metrics for Intelligent Systems Workshop, Gaithersburg, MA, 2003.

[19] Olsen, D. R. and Wood, S. B. Fan-out: measuring human control of multiple robots. Proceedings of the Conference on Human Factors in Computing Systems (CHI), 2004.

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

[21] Sidner, C. L., et al. The effect of head-nod recognition in human-robot conversation. HRI2006, pp. 290-296, 2006.

[22] Siegwart, R. et al. Robox at expo.02: A Large Scale Installation of Personal Robots, Robotics and Autonomous Systems, 42(3), pp. 203-222, 2003.

[23] Shiomi, M., et al. A Semi-autonomous Communication Robot-A Field Trial at a Train Station-. HRI2008, pp. 303-310, 2008.

[24] Shiomi, M., et al. Interactive Humanoid Robots for a Science Museum. HRI2006, pp. 305-312, 2006.

[25] Shiwa, T. et al. How quickly should communication robots respond? HRI2008, pp.153-160, 2008.

[26] Thrun, S. et al. Minerva: A second-generation museum tour-guide robot. ICRA1999, pp. 1999–2005, 1999.

[27] Walters, M. L. et al. Human approach distances to a mechanical-looking robot with different robot voice styles. IEEE Int. Symposium on Robot and Human Interactive Communication (Ro-Man 2008), pp. 707-712, 2008.