Dylan F. Glas, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita, Temporal Awareness in Teleoperation of Conversational Robots, in *IEEE Transactions on Systems, Man, and Cybernetics -- Part A: Systems and Humans*. Vol. 42, No. 4, pp. 905-919, July 2012.
This is an author's preprint version. The final manuscript is available at DOI: 10.1109/TSMCA.2011.2181162

Temporal Awareness in Teleoperation of Conversational Robots

Dylan F. Glas, Takayuki Kanda, Hiroshi Ishiguro, *Member, IEEE*, and Norihiro Hagita, *Senior Member, IEEE*

***Abstract*— Awareness of time is particularly important for teleoperation of conversational robots, both for controlling the robot and for estimating interaction success, because people have a low tolerance for long pauses in conversation. Findings have shown that people engaged in high-workload tasks tend to underestimate the passage of time. This study confirms that this problem exists for operators controlling a conversational robot, and it investigates mechanisms for improving temporal awareness and task performance while minimizing workload.**

In a laboratory experiment, two approaches to helping an operator perform various information input tasks were compared: first, assisting temporal awareness by using a clock display, and second, using autonomy to assist one of the operator's time-dependent tasks. Results revealed that assisting the task itself, even without the clock, improved not only task performance but also the operator's temporal awareness. However, results regarding the effect of the clock were ambiguous: it increased workload in general and did not help temporal awareness overall, but it did improve temporal awareness for text-entry tasks in particular.

As text-entry is an important task for teleoperation of social robots, we further investigated the problem of improving temporal awareness during text-entry tasks. As the first experiment suggested the effectiveness of a clock, we further validated that the clock is specifically useful to improve temporal awareness. These results showed that the clock did not increase workload for text-entry tasks; however, for touch-typing operators, the results suggested that showing a clock after the end of an interaction, rather than continuously throughout the task, could lower the operator's perceived workload.

***Index Terms*—situation awareness, social robots, teleoperation, time estimation**

Manuscript received June 7, 2011.

D. F. Glas is with the Intelligent Robotics and Communication Laboratories at ATR (Advanced Telecommunications Research Institute International), Kyoto 619-0288, Japan. (Corresponding author, phone: +81-774-95-1405; fax: +81-774-95-1408; e-mail: dylan@atr.jp).

T. Kanda is with the Intelligent Robotics and Communication Laboratories at ATR. (e-mail: kanda@atr.jp).

H. Ishiguro is with the Intelligent Robotics Laboratory at the Graduate School of Engineering Science at Osaka University, Osaka 565-0871, Japan, and the Intelligent Robotics and Communication Laboratories at ATR. (e-mail: ishiguro@sys.es.osaka-u.ac.jp).

N. Hagita is with the Intelligent Robotics and Communication Laboratories at ATR. (e-mail: hagita@atr.jp).

This work was supported by the Ministry of Internal Affairs and Communications of Japan.

I. Introduction

SOCIAL robots operating in field environments face recognition challenges far beyond the abilities of today's autonomy, and some level of teleoperation is necessary to support conversation. We have found that the highly time-sensitive nature of conversation presents unique challenges in teleoperation, particularly regarding the awareness of time.

Fig. 1. Robovie gives directions to customers in a busy shopping center.

Consider these two anecdotal reports from a field trial we recently conducted, in which an operator simultaneously controlled the conversations of four robots conversing with customers in a shopping center (Fig. 1):

Operator: *The operator sat tensely in the control booth, watching the flashing robot status indicators. Gripping the mouse tightly, he scrambled to find destinations on maps and choose robot commands from menus, punctuating the intense silence with frustrated outbursts, "Gaaa! No! Wait!"*
Finally he emerged from the control booth, exhausted from the ordeal but yet grinning, like an athlete walking off the field after a hard-won victory. "I did it!" he exclaimed. "I think I might even be able to handle 5 robots!"

Customer: *We spoke with one of the customers after he had interacted with one of the robots. "It was very disappointing," he said. "The robot didn't seem to listen to me. I stood there for almost a minute before it finally answered my question."*

The magnitude of this disconnect was perplexing. The customer considered the interaction a failure, while the operator believed he had been successful. How could the operator fail to understand how long the customer had been waiting?

Situation awareness is an important focus of many studies in the field of robot teleoperation. It is common for robot operators deeply engaged in a task to develop "tunnel vision," a condition in which awareness is highly focused, and the operator loses the ability to monitor background information. In studies of robots for navigation, search, and manipulation, situation awareness is usually considered in terms of the perception of spatial phenomena [1]. Many studies measure the amount of time required to gain situation awareness, but few address the direct awareness of time itself, *i.e.* time estimation. However, we found awareness of time to be an important consideration for conversational robots. We believe that in our case, the operator's "tunnel vision" was focused on the immediate informational tasks, and that the operator consequently lost awareness of the customer's situation and the passage of time.

Robot operators have many tasks to perform. They may need to type in sentences, search for information, send commands to control gestures and speech, or use "conversation fillers" [2] to stall for time while performing these tasks. With such a high workload, we have seen the operators of such systems lose awareness of the passage of time. Thus not only might an operator with a heavy workload make a customer wait for an excessive period of time, but in our experience the operator is often *not even aware of this fact!*

In this paper we address the phenomenon of temporal awareness loss in teleoperation of conversational robots. We show experimental results confirming that operators under high workload underestimate the passage of time during operation. We then propose two mechanisms for addressing the problem and evaluate them with respect to situation awareness, perceived workload, and overall effectiveness.

II. Related Work

In psychology and cognitive science, *time estimation* has been studied. In the context of teleoperation studies in HRI, *situation awareness* is considered to be important, but mainly about spatial situations. It has also been shown that *shared autonomy* can be a great help if well prepared. In social robots, the issue of *timing* has been found to be important. Here, we summarize literature in three domains: *time estimation, situation awareness and shared autonomy in HRI,* and *timing in social robots*, all of which come together in this study.

A. Time Estimation

Literature in psychology and cognitive science has revealed how people's sense of time varies. First, they have found that perception of short time, ranging from 30 ms to a few seconds (between 1 second [3] and 5 seconds [4]), and perception of long time are different problems. The former is called *time perception*, and the latter is called *time estimation*. In the context of our study, since each operation of conversational robot usually takes more than a few seconds, we are interested in *time estimation*. In addition, our study is concerned with the case where a person knows that they need to estimate time, which is categorized as *prospective* time estimation in the literature, in opposed to the case where a person is only asked afterwards, called *retrospective* time estimation [4].

For the *prospective time estimation* problem, the literature is in agreement that busy people estimate time as being shorter than the actual elapsed time. For instance, it was found that the passage of the time is estimated to be shorter when a person is engaged in a concurrent task in addition to the time estimation task, and when the concurrent task is interesting and complex [4]. Devoting more attention to non-temporal events and having a higher information processing load result in shorter time estimation [5]. Having greater demands on short-term-memory also results in shorter time estimation [6], [7], [8]. Researchers have started to integrate previous theories into a cognitive architecture [9].

B. Situation awareness and shared autonomy in HRI

In studies of teleoperation of robots for navigation and finding targets, the importance of *situation awareness* has been demonstrated [1]. Various methods have been developed to assist an operator's situation awareness, such as visualization of directions [10], maps [11], [12], and surrounding scenes [13].

While these studies address situation awareness for spatial information, to our knowledge few studies have addressed the awareness of the passage of time, that is, the problem of time estimation. In contrast, teleoperation of social robots is highly time-critical [14],[15]. This does not simply mean that an operator needs to make quick decisions; instead, the operator needs to make appropriate decisions based on time estimation. Note that previous studies considered the importance of time, but only as a metric, *e.g.* temporal demand [16], and efficiency measured by time [17], not as a problem of operator perception during operation.

In a study of *shared autonomy* (adjustable/sliding autonomy*)*, researchers have found that autonomy can help operators. For instance, autonomy was used for supporting navigation and manipulation by replaying scenes in the past [18], and for alerting about obstacles and helping with path planning [19]. Strategies for shared autonomy have also been studied. For instance, Hardin & Goodrich found that a mixed-initiative strategy performed better than adjustable and adaptive autonomy in search and rescue tasks [20].

C. Timing in social robots

When a customer is interacting with a teleoperated robot, the customer is engaged in a face-to-face interaction; however, the operator is engaged in information-management tasks using a graphical computer interface. Studies have shown that computer-mediated communication has different temporal qualities from face-to-face communication [21], suggesting that there may be an imbalance between the customer's temporal context and the operator's temporal context. This disconnect could prevent the operator from relying on an intuitive sense of the flow of time during the conversation.

Recent studies in social robots have started to highlight the importance of timing. In human communication, there is a pause during turn-taking [22],[23]. The length of the pause ranges from 620 to 770 ms [24]. In human-robot interaction, such natural pauses in human communication have been replicated [25]. Robins *et al*. explored how different response times change user reactions to a robot in a setting where a child and a robot are playing drums together [26]. It is reported that people sometimes prefer longer pauses, *e.g.* in the case where a robot is providing route directions, when people need to process information extensively [27].

One of the important related works is a study about *conversational fillers*. Shiwa *et al.* considered the problem of moderating people's negative feelings when a robot cannot make a quick response within a second. They demonstrated that such conversational fillers as "*etto*" can help a robot comfortably placate a user when it cannot respond immediately [2]. This technique has already been used in teleoperation of social robots in a field trial to moderate customers' frustration towards slow responses [28].

III. Problem Verification

In conversation, we rely on our time-estimation abilities and intuition to manage timing. If an operator has a distorted sense of the passage of time, it follows that we cannot rely on that operator's intuition to manage the timing of the interaction. Errors in time estimation can lead to awkward interactions, excessive wait times, inappropriate utterances, and a false perception of task success.

A. Experimental verification

We performed a laboratory experiment to verify whether this distortion of temporal awareness can be shown to occur in teleoperation of conversational robots.

As the literature suggests that having higher workload (*e.g.* information processing load or short-term-memory demand) results in shorter time estimation [4]-[8], we hypothesized that the operator would underestimate the amount of time that had passed, and that the magnitude of this error would increase with the operator's workload.

1) Experimental Setup

For this experiment, a computer functioning as a teleoperation console was placed in one room, and a robot was placed in another. In a camera shop scenario, participants controlled the robot to answer questions from an experimenter about different models of digital cameras.

12 undergraduate, native Japanese speakers (5 female and 7 male, average age 20.8, standard deviation 2.05 years) participated in this study, for which they were paid. None had any experience teleoperating our robots.

a) Robot

For all of our experiments, Robovie II humanoid robots were used, as shown in Fig. 2. Robovie II is capable of humanlike expressions with a 3-DOF (degrees of freedom) head, 4-DOF arms, and 2-DOF eye cameras. It can gesture and perform speech synthesis according to commands sent from a teleoperation system, and it can stream video and audio to a remote operator.

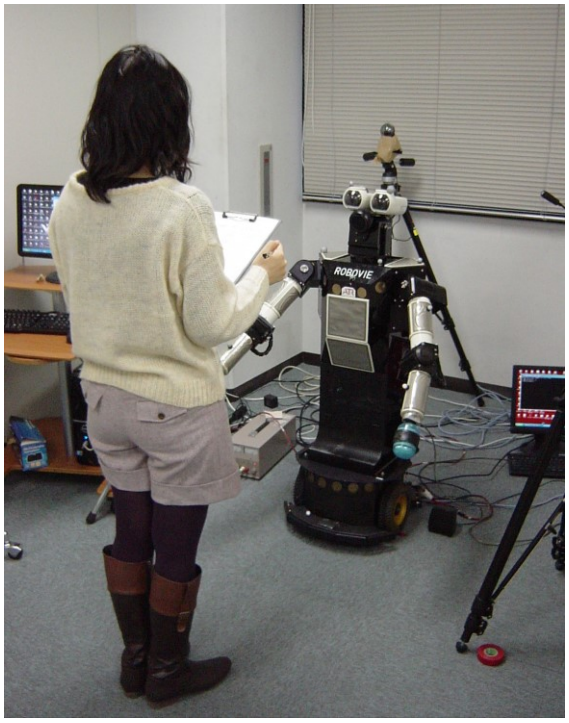


Fig. 2. Robovie II, the communication robot used in our experiments.

b) Teleoperation interface

The teleoperation interface used for this experiment was a Java application showing a video feed from the robot's camera at the top of the screen, and a control panel for the operator in the lower part of the screen. The control panel was very simple, with only two buttons available to the operator at any time.

1) Procedure

Each participant controlled the robot for six interactions, two for each of the three workload conditions (low, medium, and high). The order of these conditions was counterbalanced.

For each question, the operator was presented with a choice of two category buttons. After choosing one of the options, the operator was faced with another binary choice, continuing until the end of the tree, where the operator could choose one of two utterances for the robot to speak. This binary tree design enabled the workload of the task to be controlled precisely by adjusting the depth of the tree. In this way, we were able to create low, medium, and high workload conditions, using 1, 3, and 6 choices respectively. Fig. 3 shows an example of our interface. For workload consistency, operators were instructed to continue choosing the categories that seemed most appropriate, even after making a mistake.

After each interaction, the operator recorded an estimate of the absolute number of seconds which had elapsed between the asking of the question and the operator's response.

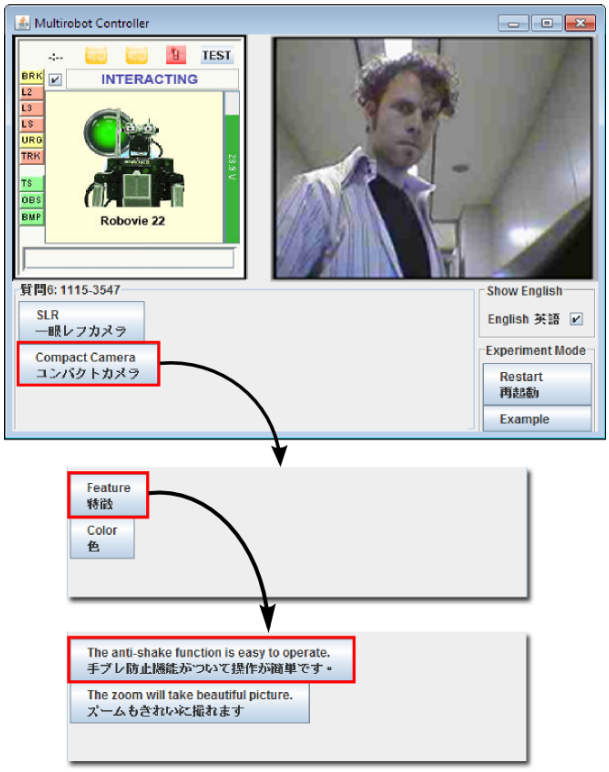


Fig. 3. Example of the "binary tree" interface to answer the question, "does this compact camera have shake reduction?"

2) Results

As we had predicted, the participants underestimated the amount of elapsed time when workload was high. Fig. 4 compares the average operation time for each condition with the average time estimated by the operators. Note that each participant performed two tasks for each condition, so we took the average of two measurements for each condition.

As these results show, the operators slightly *over*estimated the time by 1.2 seconds in the low-workload case, and they underestimated it by 1.3 seconds in the medium workload case, and by 7.7 seconds in the high-workload case.

For this time-estimation gap (*i.e.* real time minus estimated time), a repeated-measures ANOVA (Analysis of variance) was conducted with one within-subject factor, workload. The Huynh-Feldt ε correction was used to evaluate F ratios for repeated measures. A significant main effect was found ($F(2, 26)=22.790$, $p<.001$, $\varepsilon=.772$, partial $\eta^2=.637$).

A multiple comparison with the Bonferroni method was conducted for the workload factor, revealing significant differences among all pairs ($p<.001$ for comparison of the high-low pair, $p<.01$ for the medium-low pair, and $p<.05$ for the high-medium pair).

1) Discussion

These results clearly show the phenomenon with which we are concerned: operators tend to underestimate the passage of time in high-workload conditions, sometimes dramatically. In a real conversational interaction, this phenomenon could result in an operator making a customer wait for an unreasonably long time, without even realizing how much time was passing.

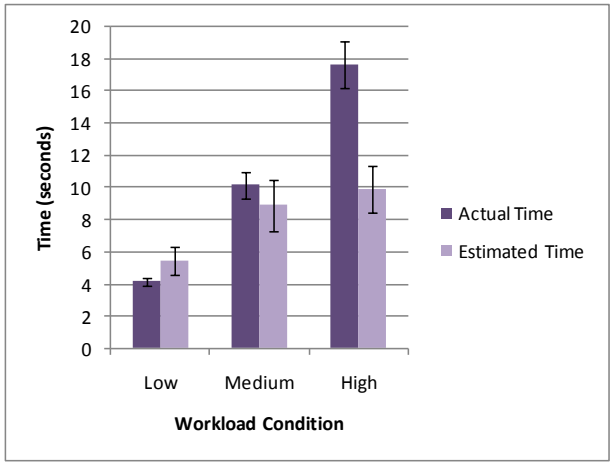


Fig. 4. Comparison of operator time estimates with actual elapsed time for three workload conditions.

The participants in this experiment were not experienced robot operators, and it is probable that time estimation could be improved through training. However, it is our hope that the problem can be addressed through user interface design to allow a wide range of operators to control robots without extensive and specialized training, and as our field trial experiences show, even expert operators experience this phenomenon to some degree.

IV. Techniques For Assisting Teleoperation

Having verified the problem, we next examined the basic tasks necessary for conversational teleoperation, and we developed two techniques to help mitigate the problem of impaired temporal awareness.

A. Teleoperation Task

In real-world teleoperation situations, it is necessary to maintain a customer's attention when an operator is unable to respond quickly. For this purpose, we often use *conversational fillers*. These are interjections such as "hmm" which provide some feedback to the customer until the operator can provide a proper response. The study in [2] demonstrates that the appropriate use of conversational fillers improves customer satisfaction.

In our field trials, the operators manually actuate these conversational fillers in addition to operating the other controls in the interface. If an utterance will take a long time to type, the operator will click a button to start a filler before typing, with the goal of keeping the silence time low, *e.g.* below 5 seconds.

The operator is thus responsible for two main tasks with different temporal awareness requirements. The first is selecting or typing appropriate utterances. For this task, overall awareness of whether or not a customer has been made to wait too long may influence the operator's choice of utterances. Temporal awareness on this scale is our primary concern.

The second task is actuating conversational fillers when necessary. This task requires a more precise awareness of the amount of time that has passed, and task success is sensitive to small time estimation errors.

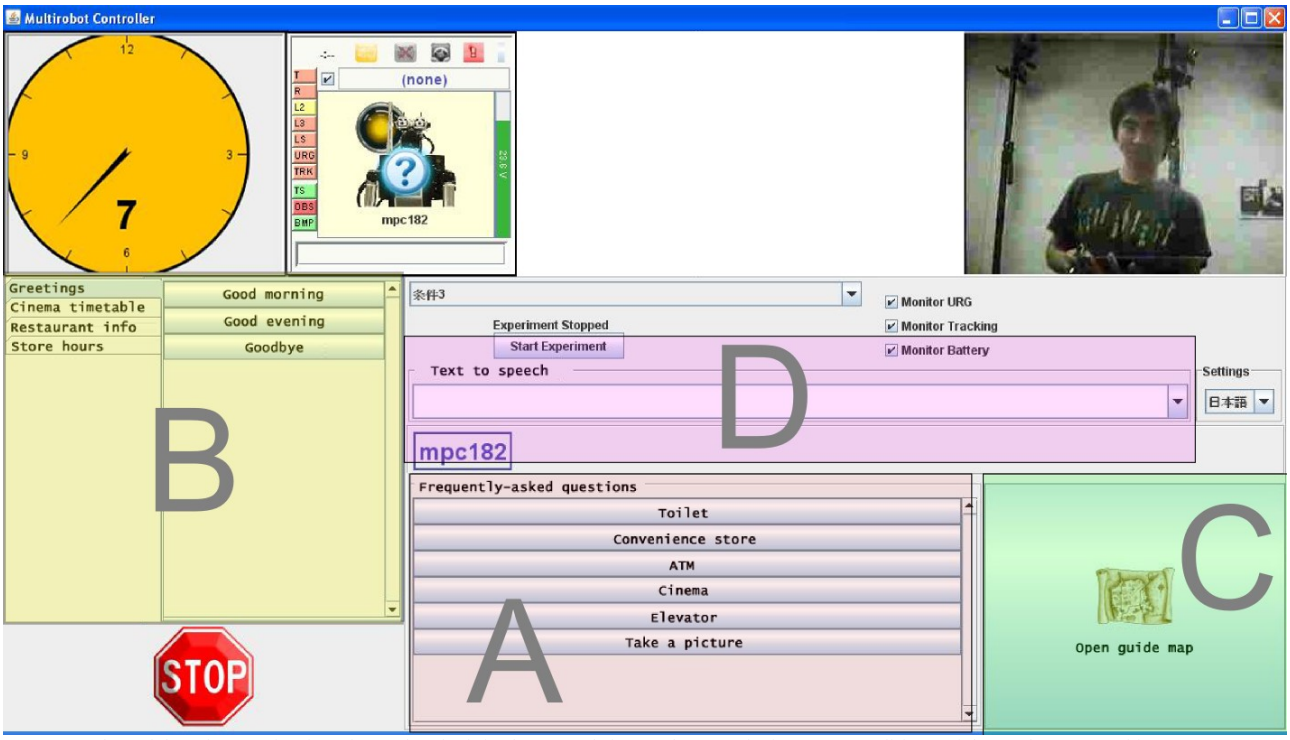


Fig. 5. Screenshot of the teleoperation interface. The clock in the upper left was shown only for some conditions of the study presented in Sec. VI.

B. *Assisting Operator Awareness*

The first technique we evaluated was using a clock display to explicitly assist the temporal awareness of the operator. We chose to use a clock (shown in Fig. 5) which displays time through a rotating second hand and a digital display of seconds.

Our hypothesis was that this mechanism would improve the operator's time estimation. We expected it would also help the operator use conversational fillers more effectively, which should reduce the number of long silences. However, as it does not change the operator's task, we predicted that it would not reduce the operator's workload or improve overall response time.

C. *Automating Conversational Fillers*

The second approach we evaluated was the automation of conversational fillers to simplify the operator's task.

Note that different kinds of conversational fillers are appropriate for short and long pauses, and a conversational filler should not be used if the operator is expected to respond quickly. We developed a simple model to predict the operator's response time, and used this prediction to make decisions about the timing and usage of conversational fillers.

Our hypothesis was that this would improve the timing of conversational fillers and reduce long silences, which is assumed to improve customer satisfaction. We also hypothesized that this mechanism would reduce the operator's workload and improve the operator's response time, as it simplifies the operator's task. We did not predict that it would necessarily improve the operator's estimation of time.

V. Estimating Operation Time

To model the amount of time required by the operator to respond to a question from the customer, we will consider the operator's *response time* to be the sum of the operator's *thinking time* and *actuation time*. For robots providing simple services, we assume that the majority of inquiries will be simple, factual questions, for which *thinking time* can be approximated as being constant.

Next, we model *actuation time* as being a function of the type of input task being performed by the operator, such as entering a phrase, or finding a place on a map. From our field trial experiences, we have observed that this is often the case, as text entry and map selection tasks take much longer than simply clicking a button or choosing an option from a menu.

While recognizing that many factors, such as training time and computer experience, can affect individual response times, we performed a study to generate a basic model to predict the amount of time required by college-age, first-time operators using our interface to respond to a set of predefined questions.

A. *Objective*

The objective of this study was to create a simple empirical model enabling us to predict operation times for an operator using our interface, based on the input task being performed. The four input tasks we investigated were as follows:

- Simple choice: Clicking a single button
- Categorized choice: Choosing an item from a tabbed menu
- Find a place: Choosing a location from a map
- Enter a phrase: Direct text entry via the keyboard

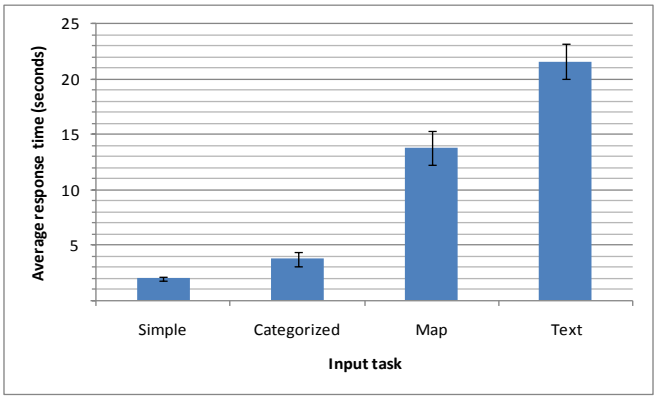


Fig. 6. Average operator response times for four types of input tasks.

B. Setup

This study was based on a robot providing guidance and information services in a shopping mall. Each participant remotely operated the robot while one of our staff members, playing the part of a customer, asked the robot questions.

8 undergraduate, native Japanese speakers (3 female and 5 male, average age 22.5, standard deviation 1.85 years) participated in this study, for which they were paid. No participants had had prior experience operating our robots.

The teleoperation interface used for this study was based on an interface we developed for our field trials. Fig. 5 shows the graphical layout. The upper panel shows the robot's status and video from the robot's camera. The lower panel contains operator controls, featuring a fixed list of buttons in the center (area "A" in Fig. 5), a tabbed menu of buttons representing categorized behaviors on the left (area "B"), a button on the right for opening a map panel showing guide locations within the shopping mall (area "C"), and a text entry field at the top for directly entering text for the robot to speak (area "D"). The clock in the upper left was not shown during this study.

C. Procedure

Five sets of four questions (4 questions for training, and 16 questions for evaluation) were prepared, with each set including one question for each of the four input tasks. Responses to those questions were prepared for the interface.

The simple choices included answers to commonly-asked questions from our field trials, such as "where is the toilet?" and "may I take a picture?" Note that although giving directions to a location requires more than a simple utterance, it is still a closed-ended question for which a response including several gestures and utterances can be pre-programmed. Thus from an operator's perspective, responding to the question "where is the toilet?" is as simple as responding to a yes-no question such as "may I take a picture?"

The categorized responses included movie start times, sorted by movie title; restaurant recommendations, sorted by restaurant type; and shop closing times, sorted by type of shop.

For the map-based tasks, we used guide maps taken from one of our field trials, modified to show only two floors of the shopping center, and eight shops on each floor.

Finally, the text field was prepared to simulate situations that are not covered [14], *i.e.* a predefined answer for that question has not been implemented in the robot. In our field trials, operators had the background knowledge to answer such questions. Since participants lacked such knowledge, we prepared a list of questions and answers which could not be answered using the buttons on the interface. Participants were instructed to use the text field in such cases. An example of one of these questions is, "What special events are happening this week?"

Every control in the interface was explained individually, including those inside the tabbed menus and every location on the map. The stock answers for the text entry questions were also presented. Each participant then operated the interface for four practice questions, one for each type of input task.

Participants then responded to the remaining 16 questions, asked in random order. The average response times recorded for each of the input tasks are shown in Fig. 6. Unsurprisingly, the results showed that the simple and categorized input tasks were much faster than the others, and that text entry was the slowest by far. Operator response time directly translates into customer wait time, so these values can help to predict how long a customer will be made to wait, as a function of the operator's input task.

D. Application

This model enables us to develop an automatic mechanism for inserting conversational fillers in an appropriate way.

1) Conversational filler Strategies

We developed three strategies for generating conversational fillers based on the operator's estimated response time: "no filler", "short filler", and "long filler".

No filler: According to the findings in [2], it is important for the robot to respond in some way within about two seconds. If the operator can respond in that time, no filler is required.

Short filler: If the operator is expected to take slightly longer than two seconds, a short filler is necessary. Our system uses "*etto,*" a thinking sound similar to "hmm" in English.

Long filler: For response times longer than two seconds, a long filler may be more socially appropriate than simply repeating "*etto*" several times. For long fillers, our robot says different phrases, like "*chotto matte ne*" ("please wait a moment"). The robot then continues saying fillers every 4 seconds, to signal to the customer that it is still "thinking".

1) Applying the Model

By monitoring user interface events, we can identify which input task is being performed by the operator. If the mouse pointer is detected in the fixed button panel or the tabbed menu, we assume that the operator is searching for one of the fixed or categorized choices. A click on the map button or text box indicates that the operator will use the map or enter text.

Using these detected actions and the model created here, we can make a rough prediction of the operator's response time. The predicted response time can then be used to choose the conversational filler strategy, as shown in Table I. The 7-second

demarcation between the short and long filler strategies comes from the initial filler time (2 seconds), plus the time for the filler utterance (around 1 second), and 4 seconds of silence.

VI. Experimental Comparison of Solutions

A. Experiment

A 2x2x4 within-participants factorial design was used to compare the effectiveness of these two proposed techniques. The first factor, *clock*, represents the use of the clock mechanism described in Sec. IV-B, in two levels: *clock* and *no clock*. The second factor, *filler*, represents the use of the automatic filler technique described in Sec. IV-C, in two levels: *auto-filler* and *manual-filler*. The third factor is the input task for each question, represented by the *input-task* factor in four levels: *simple*, *categorized*, *map*, and *text*.

1) Procedure

Participants operated a robot in a shopping mall scenario, using an interface like the one described in Sec. V, but with the addition of a clock display and a conversation filler button.

23 undergraduate, native Japanese speakers (15 male and 8 female, average age 21.1, standard deviation 2.0 years) participated in our experiment, for which they were paid. None had participated in the other studies in this paper.

a) Instructions

The scenario was explained to the participants, and they were shown a demonstration of a simple interaction with the robot. The robot's response time and the importance of responding quickly were discussed, and the point was repeated several times throughout the task explanation.

Every control and map location on the interface, including the clock and conversation filler button, was explained. For the *manual-filler* conditions, the operators were instructed to manually insert conversational fillers using a button on the interface, first within 2 seconds of the end of the customer's question, and afterwards never to allow more than 5 seconds of silence. They were also told to be aware of their operation time, and to estimate it after each interaction.

A four-question training session was conducted for each operator, just as in the previous study. The same list of questions from the previous study was used for this experiment.

TABLE I. Conversational Filler Strategies by Response Time

Predicted Response Time	Conversational Filler Strategy
< 2 seconds	No filler
2-7 seconds	Short filler
> 7 seconds	Long filler

b) Trials

Each trial consisted of four questions, one for each of the input tasks, which were always asked in the order: *simple*, *text*, *map*, *categorized*. Note that while the customer was asking a question, the operator's screen controls were blanked, so even if an operator could anticipate the *input-type* for the next response, no pre-actuation was possible.

Four trials were conducted for each participant, one trial for each combination of *clock* and *filler* conditions. The order of *clock* and *filler* experimental conditions was counterbalanced between participants, and question sets were also counterbalanced between conditions, to ensure that results were independent of specific question content. Each participant answered each question only once.

2) Evaluation

After each interaction, the participants estimated the time it took them to respond to that question. Then, after each trial of four questions, the participants rated their workload for the trial. For this evaluation, we used the NASA-TLX scale (Task Load Index) [16], a tool for assessing subjective workload based on six factors: mental demand, physical demand, temporal demand, operator performance, frustration, and effort.

A total of four measurements were used in this study:

- *Operation time*, from the end of the customer's question until the operator sends a command
- *Time estimation error*, calculated by subtracting the estimated time from the actual operation time
- *Silence duration*, the maximum duration of silence between robot utterances during an interaction
- *Perceived workload*, the NASA-TLX score

B. Hypotheses

To restate the hypotheses from Sec. IV in terms of the factors in this experiment, we predicted that the use of *auto-filler* would reduce *silence duration*, *perceived workload*, and *operation time*, with no effect on *time estimation error*. Furthermore, we predicted that the presence of the *clock* mechanism would reduce *time estimation error* and *silence duration*, but have no effect on *operation time* or *perceived workload*.

C. Results

The results for the four measurements are shown in Fig. 7. Full analysis is presented for all three factors (*clock, filler,* and *input-task*) for the measurements of "operation time," "silence time," and "time estimation error." Regarding "perceived workload," the NASA-TLX test was administered only once after each trial of four questions. As each trial contained all four input tasks, it was not possible to examine TLX scores for each input task separately. Hence perceived workload is analyzed here with respect to *clock* and *filler* only.

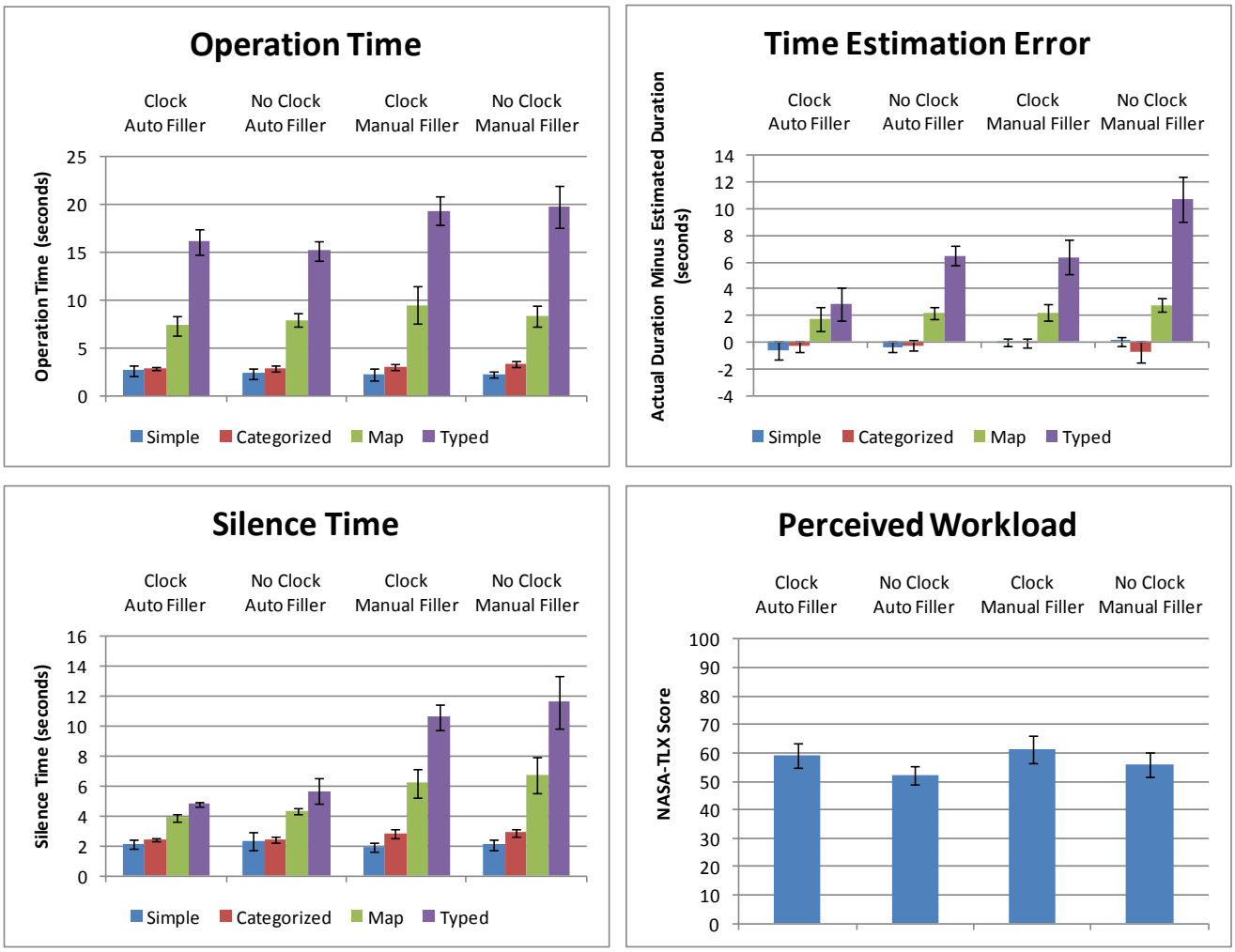


Fig. 7. Results for the four variables measured in our experiment.

filler (F(1.984, 66)=4.203, p=.022, partial η^2=.160) were significant, whereas the interaction with *clock* (F(2.316, 66)=.049, p=.967, partial η^2=.002), and the interaction among the three factors (F(2.048, 66)=.460, p=.639, partial η^2=.020) were not significant.

The interaction with filler indicates that the filler significantly reduced the operation time in *typed* input (p=.006), but the difference was not significant for the other input types: *simple* (p=.621), *categorized* (p=.204), and *map* (p=.200).

The main effect and the interaction with *filler* also indicate that operation time varied for different input tasks, as we already discovered in Sec. V. A multiple-comparison with the Bonferroni method was conducted for the four input tasks, which revealed significant differences in operation time as follows: for the *manual-filler* condition, *text* > *simple*, *categorized*, and *map* (p<.001), and *map* > *simple* and *categorized* (p<.001). There was no significant difference between *simple* and *categorized* (p=1.0). For the *auto-filler* condition, *text* > *simple*, *categorized*, and *map* (p<.001), *map* > *simple* (p<.001) and *categorized* (p=.001). There was no significant difference between *simple* and *categorized* (p=.119).

Predictions: *Auto-filler* will reduce operation time. *Clock* will not affect operation time.

within these factors (F(1,22)= .001, p=.977, partial η^2=.000).

Regarding the input-task factor, the main effect (F(1.868, 66)=33.650, p<.001, partial η^2=.605) and the interaction with *clock* (F(1.559, 66)=9.066, p=.002, partial η^2=.292), and the interaction with *filler* (F(2.384, 66)= 8.246, p<.001, partial η^2=.273) were significant, whereas the interaction among these three factors (F(1.775, 66)=.223, p=.775, partial η^2=.010) was not significant.

We analyzed the interaction with the *clock* with the Bonferroni method, which revealed that in the *text* input the *clock* effect was significant (p<.001), but no significance was found in other inputs (for *simple*: p=.496, *map*: p=.418, and *categorized*: p=.218).

We analyzed the interaction with the *filler* with the Bonferroni method, which revealed that in the *text* input the *filler* effect was significant (p<.001), and almost significant in the *map* (p=.092), but no significance was found in other inputs (for *simple*: p=.359, and *categorized*: p=.781).

Predictions: *Auto-filler* will not affect time estimation error. *Clock* will reduce time estimation error.

Results: Surprisingly, *auto-filler* significantly reduced time estimation error in *text* entry; *clock* also had the effect of reducing time estimation error in the case of *text* entry.

3) Silence duration

For maximum duration of silence, shown in Fig. 7 (lower left), a three-way repeated-measures ANOVA was conducted with three within-subject factors, *clock, filler,* and *input-task*. The Huynh-Feldt ε correction was used to evaluate F ratios for repeated measures. A significant main effect was revealed in the *filler* factor ($F(1,22)= 18.991$, $p<.001$, partial $\eta^2=.463$). No significance was found in the *clock* factor ($F(1,22)=1.433$, $p=.244$, partial $\eta^2=.061$) or in the interaction within these factors ($F(1,22)= .011$, $p=.917$, partial $\eta^2=.001$).

Regarding the input-task factor, the main effect ($F(2.173, 66)=54.385$, $p<.001$, partial $\eta^2=.712$) and the interaction with *filler* ($F(2.046, 66)=15.199$, $p<.001$, partial $\eta^2=.409$) were significant, whereas the interaction with *clock* ($F(2.703, 66)=.794$, $p=.491$, partial $\eta^2=.035$), and the interaction among these three factors ($F(3,66)=.006$, $p=.999$, partial $\eta^2=.000$) were not significant. We analyzed this significant interaction with the Bonferroni method, which revealed that in the *manual-filler* conditions, max duration of silence was longer in *categorized* ($p=.026$), *map* ($p=.007$), and *text* ($p<.001$) input, but not for *simple* input ($p=.607$). Clearly, this is because *simple* input is fast enough not to require fillers, so use of *auto-filler* did not contribute to reduce max duration of silence for *simple* input.

Predictions: *Auto-filler* and *clock* will both reduce silence duration.

Results: As predicted, the use of *auto-filler* reduced the maximum silence duration, whereas interestingly, the *clock* did not affect the maximum duration of silence, even for the manual-filler condition (in fact, a separate ANOVA was conducted only for *manual-filler* conditions which did not show any significant difference).

4) Perceived workload

For the NASA-TLX scores, shown in Fig. 7 (lower right), a two-way repeated-measures ANOVA was conducted with two within-subject factors, *clock* and *filler*. A significant main effect was revealed in clock factor ($F(1,22)= 8.204$, $p=.009$, partial $\eta^2=.272$). No significance was found in the *filler* factor ($F(1,22)=.683$, $p=.418$, partial $\eta^2=.030$) or in the interaction within these factors ($F(1,22)=.042$, $p=.840$, partial $\eta^2=.002$).

Predictions: *Auto-filler* will reduce perceived workload. *Clock* will not affect perceived workload.

Results: The presence of a *clock* increased the perceived workload, and using the *auto-filler* did not decrease the perceived workload as we had expected.

VII. Discussion and Limitations

A. Summary and interpretations

The experiment results showed that when the clock was displayed, perceived workload increased. The effect on time estimation was not significant but showed a trend ($p=.098$) that the operator had better time estimation when the clock was shown. When the automatic filler mechanism was in use, total operation time decreased, and the length of the maximum silence interval decreased. The operator's time estimation also improved, as indicated by a decrease in estimation error.

Of the four input tasks, typing was generally the most time-consuming. The analysis of interaction with the input-task factor revealed that the clock was most helpful in time estimation for the typing tasks, and auto-filler was most effective in reducing max silence duration for the typing tasks.

These results raise some questions.

Why did the clock not help time estimation so much, while auto-filler showed a clear effect? A possible explanation is that the auto-filler simplified the operator's task, resulting in better time estimation. The literature confirms that time estimation is better in less complex situations [4],[5].

The operator's task is also more complex when the clock is visible, requiring the operator to process time information in addition to other tasks. This might explain the marginal results regarding time estimation.

Another possibility is that the robot's auto-filler behavior may have provided audible feedback to the operator, although this was not the intention of its design. This feedback may have helped the operators to estimate time, since it came at regular intervals. Furthermore, the fact that the feedback came from the auditory rather than visual channel may have decreased the operator's workload, as human factors research shows that using different sensory modalities for different tasks can improve cognitive processing efficiency [29].

Why, then, did auto-filler not reduce perceived workload, even though it actually simplified the operator's task, resulting in shorter operation time? One possibility is that, as the majority of the operator's time and attention was spent on the input tasks, those tasks more strongly influenced *perceived* workload than the manual-filler task did. Yet, the fact that both operation time and time estimation were improved by using auto-filler suggests that auto-filler may in fact reduce *actual* workload.

B. Are these findings too obvious? Not to our operators.

Interestingly, we received unsolicited complaints from two of the participants who claimed the automatic filler mechanism was frustrating, because they preferred to have complete control over the system. This seems to indicate that the operators did not always perceive a need for the automatic filler mechanism, and that its benefits are not so obvious. However, in this study the automatic filler mechanism was shown to perform much better than the operator in preventing long silences.

C. Generalizability and Limitations

These findings are specific to our teleoperation system, based on four input tasks. However, these are common operational tasks for conversational robots, so the findings may be applicable to many cases of teleoperation for social robots.

We are also interested in the teleoperation of multiple robots. These findings have not been tested in that scenario. We predict that the temporal awareness problem will be more extreme with multiple robots, since the operator's task is more complex. Our solution may thus be even more effective in that case, but this remains to be tested.

VIII. Clock Display Experiment

Based on these results, we can conclude that a clock should not be shown for the input mechanisms other than text entry, since the clock does not help time estimation in those cases but does increase perceived workload.

Text entry is an important case, however, as it requires longer actuation time than the other input methods, and thus time estimation errors carry a greater risk for excessive customer wait times. For text entry, showing a clock could be useful for reducing time estimation error. It is not clear whether it would increase perceived workload, however, as our first experiment did not measure TLX scores for separate input tasks.

We thus conducted a second experiment to focus on the effect of a clock display in text-entry tasks. Our hypothesis was that the presence of a clock during text entry would improve the operator's time estimation but might also increase perceived workload by creating a feeling of time pressure.

We also evaluated an interface design in which a clock was shown only after each text entry task was complete. Our hypothesis was that showing the elapsed time after each task would increase time estimation accuracy, whereas hiding the clock during operation would reduce time pressure and thus also reduce perceived workload.

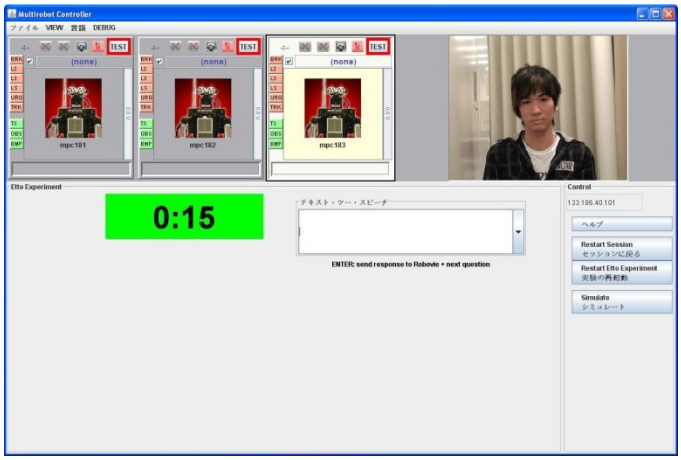


Fig. 8. User interface for the clock experiment.

Furthermore, preliminary studies showed that the effectiveness of the clock displays could be dependent upon the typing style of the operator. We observed that touch-typing operators who watched the screen while typing were more aware of time while a clock was being displayed than non-touch-typing operators who were looking at the keyboard. For this reason, we studied the effect of typing style as a factor in our experiment as well.

A. Conditions

For this experiment, we used a 2x3 between-participants factorial design with two factors: *typing style* and *clock type*.

The *typing style* factor was studied in two levels: *up-type*, meaning the operator looked at the screen while typing, and *down-type*, meaning that the operator looked at the keyboard some or all of the time while typing.

The *clock type* factor was studied in three levels: *no-clock*, *clock-during*, and *clock-after*. In the *no-clock* condition (NC), participants typed their answers to a customer's question into a text box, and no feedback was provided to them about the amount of elapsed time. In the *clock-during* condition (CD), a digital display of the number of elapsed seconds was provided on the screen while they typed the response. Finally, in the *clock-after* condition (CA), no clock was displayed while typing, but after typing was complete the display showed the total number of seconds elapsed.

Our hypotheses regarding *clock type* were as follows:

- The operator's time estimation will be improved when a clock is shown (*clock-during* and *clock-after* conditions).
- The operator's perceived workload will be higher for the *clock-during* condition than for the *clock-after* condition, because of the perceived time pressure.
- Operators will tend to type shorter utterances when a clock is present.
- Operators will tend to type faster in the *clock-during* and *clock-after* conditions.

Regarding *typing style*, we made the following hypotheses:

- *Up-type* operators will have better time estimation in the *clock-during* condition, while *down-type* operators will have better time estimation in the *clock-after* condition.
- Better time estimation will also decrease response time and utterance length, and increase perceived workload.

B. Experimental Procedure

1) Scenario

The scenario we chose for this experiment was that of an operator controlling multiple information-providing robots answering questions at a university. A total of 53 paid participants, 23 female and 30 male, took part in this experiment. All were university students (average age 20.6, standard deviation 1.7 years) and all were native Japanese speakers.

Participants performed the role of robot operator, using the interface shown in Fig. 8 to answer 15 simple questions about their university. They were told that the interface controlled multiple robots in other rooms, and that as soon as they had entered text for one robot to speak, the control would be switched to another robot. With this interface, they were instructed to answer the questions to the best of their ability based on their real experience.

Since the audio feedback from automatic conversation fillers could have been a confounding factor in the first experiment, we did not use them in the second experiment. However, as the auto-filler mechanism was shown to be useful, we assumed that such functionality would be present in a real teleoperation system and thus did not ask participants in the second experiment to enter manual fillers.

2) Questions

As reaction time was one variable of interest in this study, it was important to choose questions for which the participants would not have to look up information, but which they could answer from their background knowledge. For this reason we chose questions which we expected most students could answer about their universities, but which non-students might not know.

Three sets of 15 questions were prepared. In order to equalize difficulty between question sets, response times during preliminary trials were used to allocate questions of similar difficulty to each question set. Some examples of questions used include the following:

- Where should I go if I lose my student ID?
- How do I get to the nearest train station?
- Which courses should I take for easy A's?

For consistency of questions between trials, video and audio for all questions were recorded beforehand and then played back through the control interface during the experiments. The questions were recorded in different rooms, from the perspective of the robots' eye cameras. At least one minute of video was recorded after each question, showing the facial expressions and movements of the person waiting for the answer.

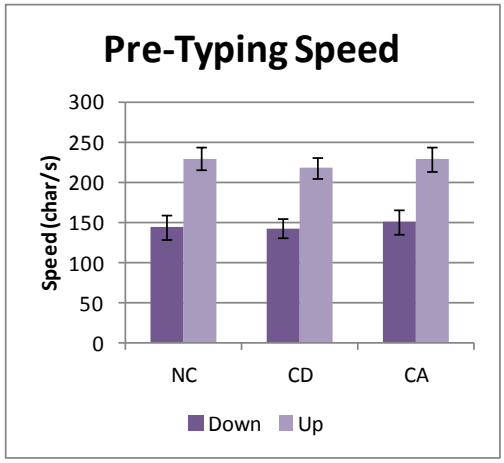


Fig. 9. Average typing speed in pre-test.

3) Procedure

Before the experiment, a typing speed test was administered to each participant. Their typing speed was recorded, and their touch-typing behavior was also observed and recorded. Participants were categorized as "up" if they looked up at the screen while typing, or as "down" if they looked down at the keyboard some or all of the time.

Within the *up-type* and *down-type* groups, participants were assigned to the different experimental conditions based on their typing speeds, with the goal of balancing typing speeds as much as possible across *clock type* conditions, as shown in Fig. 9.

The overall task was then explained to the participants. They were instructed to provide polite and complete answers to questions, but also not to make the customers wait too long. To help participants understand what a long pause would seem like to a customer, they were shown a video of a person asking questions to a robot three times. Each time, the robot paused for a different amount of time before responding. Pauses of 10, 20, and 40 seconds were shown, and the robot used conversation fillers during the pauses.

After watching the video, participants were instructed on how to use the interface, including an explanation of the clock, if one was shown. Each participant operated the interface in response to one practice question to confirm that they understood the procedure.

Participants then used the interface to answer 14 more questions in a row, all within the same *clock type* condition (*no-clock, clock-after, clock-during*). These questions measured the participants' overall performance within that condition.

Finally, participants filled out a NASA-TLX questionnaire, to evaluate their perceived workload for the task.

4) Evaluation

In summary, the following data were also collected from each participant:

- Response time for each question
- Character length of response to each question
- NASA-TLX score for each 15-question session

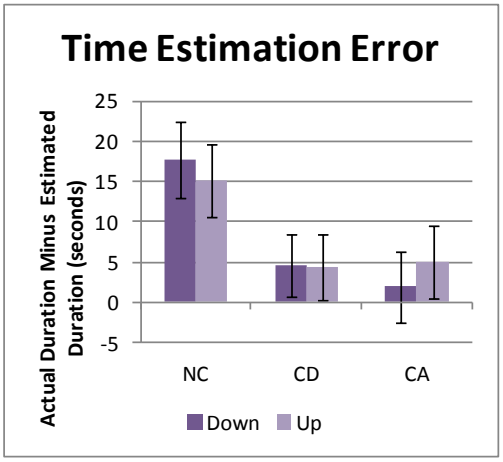


Fig. 10. Time estimation error (actual duration minus estimated duration).

C. Results

1) Time Estimation

For time estimation, we expected operators to underestimate the elapsed time in the *no-clock* (NC) condition, and to have more accurate estimation in the *clock-during* (CD) and *clock-after* (CA) conditions. We also expected *down-type* typists to have better estimation in the *clock-after* condition than in the

clock-during condition, as they spent less time looking at the screen than the *up-type* typists did. Results are shown in Fig. 10.

The time estimates of the operators were typically shorter than the actual time durations, so in this section and in Fig. 10 we will express error as "actual duration minus estimated duration," so that large values represent large errors and small values represent more accurate estimation. Thus the expression "CA<NC" indicates that the time estimation in the *clock-after* condition was more accurate than in the *no-clock* condition (the error was smaller).

A two-way ANOVA with two between-subject factors, *clock type*, and *typing style*, was conducted for time estimation. A significant main effect was revealed in *clock type* ($F(2,65)= 5.090$, $p=.009$, partial $\eta^2=.135$). Multiple comparison with the Bonferroni method revealed that there were significant differences: CA<NC ($p=.018$), CD<NC ($p=.023$), but CA=CD ($p=1.00$). No significance was found in the *typing style* factor ($F(1,65)=.001$, $p=.970$, partial $\eta^2=.000$) or in the interaction within these factors ($F(2,65)=.187$, $p=.830$, partial $\eta^2=.006$).

These results support our hypothesis that the presentation of a clock results in better time estimation. Contrary to our expectations, however, they do not show a difference between *clock-during* and *clock-after* based on *typing style*. This may be due to the fact that *down-type* operators do look at the screen from time to time to confirm they have typed the correct phrase.

2) Perceived workload

Our expectation was that the presence of a clock would increase the operator's perceived time pressure and that this would be measurable by a NASA-TLX evaluation of perceived workload, shown in Fig. 11. Furthermore, we expected that *up-type* operators would perceive higher workload in the *clock-during* condition, whereas *down-type* operators would perceive higher workload in the *clock-after* condition.

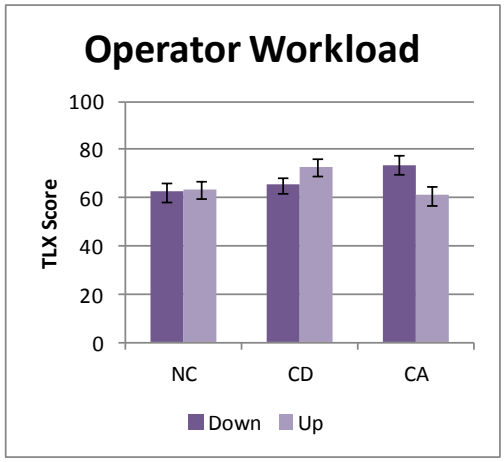


Fig. 11. Perceived workload, as measured by NASA-TLX score.

A two-way ANOVA with two between-subject factors, *clock type* and *typing style*, was conducted for TLX score. There were no significance in the *clock type* factor ($F(2,65)=1.417$, $p=.250$, partial $\eta^2=.042$) or *typing style* factor ($F(1,65)=.204$, $p=.653$, partial $\eta^2=.003$), but the interaction within these factors was significant ($F(2,65)= 4.023$, $p=.023$, partial $\eta^2=.110$).

We analyzed this significant interaction with the Bonferroni method, which revealed that in the *up-type* condition there is an almost-significant difference between CA and CD ($p=.074$), but no significance found in other inputs (for up-type, CA-NC: $p=1.000$, CD-NC: $p=.228$; for down-type, CA-NC: $p=.140$, CA-CD: $p=.299$, CD-NC: $p=1.000$).

These results suggest that for *up-type* operators, the *clock-after* method may be better in terms of reducing the operator's perceived workload. Note that for most applications, it is likely that touch-typists will be employed as operators.

3) Response Time and Character Length

We expected that both response time and character length would be lower for the *clock-during* and *clock-after* conditions, compared with *no-clock*, and that both would be lower for *up-type* operation than for *down-type*. The results for these two measurements are shown in Fig. 12.

For response time, we conducted a two-way ANOVA with two between-subject factors, *clock type* and *typing style*. No significance was found in the *clock type* factor ($F(2,65)=.109$, $p=.897$, partial $\eta^2=.003$), in the *typing style* factor ($F(1,65)=.088$, $p=.768$, partial $\eta^2=.001$) or in the interaction within these factors ($F(2,65)=.258$, $p=.773$, partial $\eta^2=.008$).

For character length, we also conducted a two-way ANOVA with two between-subject factors, *clock type* and *typing style*. In this case, a significant main effect was revealed in the *typing style* factor ($F(1,65)= 14.423$, $p=.000$, partial $\eta^2=.182$). No significance was found in the *clock type* factor ($F(2,65)=.292$, $p=.748$, partial $\eta^2=.009$) or in the interaction within these factors ($F(2,65)=.462$, $p=.632$, partial $\eta^2=.014$).

Interestingly, these results do not show any significant difference in response time or character length based on *clock-type*, contrary to our expectations.

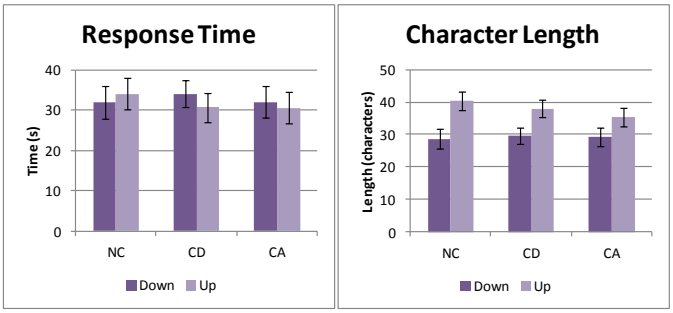


Fig. 12. Average response time, in seconds, and average character length of responses, in Japanese characters.

The results do show that although response time did not vary with *typing style*, the *up-type* typists provided longer (and ostensibly better) responses. We attribute this to the fact that the *up-type* operators had a higher average typing speed.

D. Discussion of Results

In our previous experiment, we observed a trade-off, in which the display of a clock improved the operator's time estimation, but at the cost of an increase in perceived workload.

In this experiment, we observed that the clock display again improved the operator's time estimation, but this time the effect on perceived workload was not as evident. As this experiment only examined text entry, it is possible that the observed effect in the previous experiment mainly occurred in the other, faster, input methods.

Our results suggested that, for the case of *up-type* operators, showing the clock after the entry of each utterance (*clock-after*) resulted in lower perceived workload than showing the clock throughout the task of text entry (*clock-during*). This trend did not show strong significance (p=.074), but the results suggest that the *clock-after* technique might be useful for improving the time estimation of touch-typing operators without increasing their perceived workload.

We did not see a direct association between time estimation and operator performance results, either in silence time in the first experiment, or in response time in the second experiment.

We interpret this data by considering the relationship between *time estimation* and *time pressure*. The average response time was around 32 seconds. Even underestimating this time by 15 seconds, an operator would still believe that it took 17 seconds to type the response. Yet in normal conversation, a person would respond to this question far more quickly, probably within 2 or 3 seconds. It is possible that there is little difference in the *time pressure* an operator feels after 17 seconds or 27 seconds, as both of these times are far beyond what could be considered a "normal" human response time.

That would explain why little difference in response times is visible between *clock type* conditions. However, if the operators really do feel the same amount of time pressure, then why would there be a difference in character length of the operators' responses between *typing style* conditions? We believe that this could be due to touch-typists being more fluent with keyboard entry and thus accustomed to entering longer text. If slower typists are less familiar with using a keyboard, they might naturally enter shorter, less complete answers. Thus, in applications where the quality of an answer is important, we might expect faster typists to provide *better* answers than slower typists, not simply the same answers in a shorter time, although response quality was not evaluated in this study.

IX. Discussion

Temporal awareness is important in teleoperation of conversational robots both in an immediate sense, because people have a low tolerance for long pauses in conversation, and in an overall sense, because understanding how long a customer has been waiting is important in choosing what to say. Thus impaired temporal awareness affects both utterance timing and the content of the conversation itself.

A. Partial Autonomy

In this study, partial autonomy was used to help simplify the operator's task. As technology progresses, more autonomy will become feasible. Will progress in such a direction eliminate the problem of temporal awareness? We believe it will not.

In future systems, we assume that many aspects of dialog management such as turn-taking [30] will be automated. Simple control tasks will be handled autonomously, and an operator will be responsible for handling complex, exceptional tasks that cannot be automated. As this autonomy improves over time, one operator will be able to control more and more robots.

Thus in future systems, we expect the operator's tasks to be more complex and less routine. The operator may spend less time performing direct control of minor utterances, focusing instead on high-level decisions and complex utterances. In this sense, temporal awareness will become less important in terms of immediate utterance timing, and more important in terms of choosing appropriate things for the robot to say.

B. Interaction Asymmetries

As mentioned earlier, one possible reason for the operator's poor temporal awareness is the asymmetry of the interaction, and reducing this asymmetry could help moderate the temporal awareness problem. There are two parts to be considered in this asymmetry: the task and the modality.

In terms of the task, the operator is entering data or looking up information in a map or a database, while the customer is asking a robot for information. In future systems, as the operator's task complexity increases, we expect that the task asymmetry will also increase. Both the increased complexity and the increased asymmetry may contribute to impaired temporal awareness.

In terms of the modality, the operator is interacting with a graphical computer interface, while the customer is face-to-face with a physical robot. To reduce the severity of this asymmetry, an immersive telepresence approach might help. Combining natural gesture control, as in [31] and [32], with an immersive first-person video feed [33] could reduce this asymmetry and provide the operator with a more natural sense of participating in a face-to-face interaction.

C. Limitations of this study

1) Customer experience

The experience of the customer interacting with the robot was not analyzed here, and the operator's performance was only examined numerically. The significance of an operator's temporal awareness as it affects the overall customer's experience is not easy to measure directly, and the importance of appropriate timing might be dependent upon the conversational context or other social factors. Considering the customer as a human element, there might be social ways to mitigate the sensitivity of customers to wait time.

2) Interaction complexity

Another limitation is that the interactions used in this study were simple question-and-answer exchanges. While other dialogue patterns are certainly possible, we believe that question-and-answer interactions will be quite common, particularly in the service robot domain, where interactive robots will often be providing information to people.

Another point that must be considered is that many human-robot interactions will likely extend beyond "single-round" exchanges.

Our current study addresses the problem where an operator performs one input task per interaction; however, as shown in the introduction, the temporal awareness problem becomes more serious when an operator is continuously busy with many tasks. We think it likely that an operator's small judgment errors due to inaccurate temporal awareness may accumulate over several rounds of a conversation to cause significant frustration to a customer.

3) Multiple Robot Control

While we have seen that temporal awareness is an issue even when controlling one robot, the original problem presented in the introduction was a case of multiple-robot control, which presents new challenges. Multitasking in general has been shown to impair temporal awareness. Additionally, to enable an operator to focus on one conversation at a time, auditory information from other robots would need to be selectively muted. In such a case, the operator would be even less aware of wait time for robots that were not currently the focus of attention, and explicit mechanisms might be necessary for communicating this wait time.

4) Social Feedback

Another interesting issue that was raised during the final study in this paper was the effect of the operator seeing video of the customer. While some participants ignored the video feed while typing their responses, others indicated that they felt pressured by seeing the facial expressions of the impatient customer. If such social cues can be transmitted effectively, then it is possible that the operator's temporal context might more closely approximate that of a face-to-face conversation.

X. Conclusions

In this study, we have empirically demonstrated that the time estimation ability of operators controlling conversational robots can be impaired under high workload conditions. We have also conducted a comparison of two approaches to addressing this problem: by providing temporal information explicitly through a clock display, and by using autonomy to reduce the operator's task load. The results showed that the clock display alone did not significantly improve performance, but that it did increase the operator's perceived workload. The partial autonomy resulted in better performance as well as improved temporal awareness, without significantly affecting perceived workload.

Next, we examined the effectiveness of the clock display for text entry in particular, and found that while the clock displays significantly improved time estimation, we did not see a significant influence on the length of typed responses. The results also showed an almost-significant trend among touch-typists in which showing a clock after the finish of each operation resulted in a lower perceived workload than showing a clock throughout operation, although these two conditions yielded the same improvement in temporal awareness.

An interesting conclusion of this study is that indirectly supporting temporal awareness by simplifying an operator's task may be better in some cases than direct support, as our first experiment found perceived workload to be lower when the clock was not visible. This suggests that if better awareness can be achieved by reducing the operator's task complexity, then withholding information from the operator might be beneficial.

Finally, these findings are complemented by the technical contribution of our successful implementation of an automatic filler mechanism. Our simple approach of inferring the input task from mouse movements worked well for the tasks in this study, in that it limited silence time much more effectively than manual control. This technique was not always successful, however, and the operation task was not predicted accurately every time. For higher accuracy, it may be possible to incorporate information from interaction context or history to predict the operation task, and to extend the timing model to incorporate thinking time as well as actuation time.

Acknowledgment

We would like to thank Phoebe Liu for all the hard work that made these experiments possible.

References

[1] J. Drury, L. Riek, and N. Rackliffe, "A decomposition of UAV-related situation awareness," *1st ACM/IEEE Interational Conference on Human-Robot Interaction*, pp. 88-94, 2006.

[2] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro, N. Hagita, "How Quickly Should a Communication Robot Respond?," *International Journal of Social Robotics*, 1(2), pp. 141-155, 2009.

[3] P. A. Lewis and R. C. Miall, "Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging, current opinion in neurobiology," *Current Opinion in Neurobiology*, vol. 13, pp.250-255, 2003.

[4] P. Fraisse, "Perception and estimation of time", *Annual Review of Psychology*, vol. 35, pp. 1-36, 1984.

[5] R. E. Hicks, G. W. Miller, G. Gaes, and K. Bierman, "Concurrent processing demands and the experience of time-in-passing", *American Journal of Psychology*, vol. 90, pp. 431-446, 1977.

[6] C. Fortin and R. Rousseau, "Time estimation as an index of processing demand in memory search", *Perception & Psychophysics*, vol. 42, pp. 377-382, 1987.

[7] C. Fortin, R. Rousseau, P. Bourque, and E. Krouac, "Time estimation and concurrent nontemporal processing: specific interference from short-term-memory demands," *Perception & Psychophysics*, vol 53, pp. 536-548, 1993.

[8] C. Fortin and R. Breton, "Temporal interval production and processing in working memory", *Perception and Psychophysics*, vol. 57, pp. 203-215, 1995.

[9] N. A. Taatgen, H. Rijn, and J. Anderson, "An integrated theory of prospective time interval estimation: the role of cognition, attention, and learning," *Psychological Review*, vol. 114, no. 3, pp. 577-598, 2007.

[10] C. Humphrey and J. Adams, "Compass Visualizations for Human-Robotic Interaction," *3rd ACM/IEEE International Conference on Human-Robot Interaction*, pp. 49-56, 2008.

[11] C. Nielsen and M. Goodrich, "Comparing the usefulness of video and map information in navigation tasks," *1st ACM/IEEE International Conference on Human-Robot Interaction*, pp. 95-101, 2006.

[12] C. Nielsen, M. Goodrich, and R. Ricks. "Ecological interfaces for improving mobile robot teleoperation," *IEEE Transactions on Robotics*, vol. 23(5), pp. 927–941, October 2007.

[13] F. Ferland, F. Pomerleau, C. Dinh, and F. Michaud, "Egocentric and Exocentric Teleoperation Interface using Real-time, 3D Video Projection," *4th ACM/IEEE Internatinal Conference on Human-Robot Interaction*, pp. 37-44, 2009.

[14] D. F. Glas, T. Kanda, H. Ishiguro and N. Hagita, "Simultaneous Teleoperation of Multiple Social Robots," *3rd ACM/IEEE Conference on Human-Robot Interaction* , pp.311-318, 2008.

[15] D. F. Glas, T. Kanda, H. Ishiguro and N. Hagita, **"**Field Trial for Simultaneous Teleoperation of Mobile Social Robots**,"** *4th ACM/IEEE Conference on Human-Robot Interaction*, pp.149-156, 2009.

[16] S.G. Hart and L.E. Staveland, "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research," In *Human Mental Workload*, P. Hancock, N. Mesh-kati (Eds.) pp. 139 – 183, 1988.

[17] A. Steinfeld, *et al*., "Common metrics for human-robot interaction," *1st ACM/IEEE International Conference on Human-Robot Interaction*, pp. 33-40, 2006.

[18] B. Sellner, L. Hiatt, R. Simmons, and S. Singh, "Attaining situational awareness for sliding autonomy," *1st ACM/IEEE International Conference on Human-Robot Interaction*, pp. 80-87, 2006.

[19] M. Goodrich, T. McLain, J. Anderson, J. Sun, and J. Crandall, "Managing autonomy in robot teams: observations from four experiments," *2nd ACM/IEEE International Conference on Human-Robot Interaction*, pp. 25-32, 2007.

[20] B. Hardin and M. A. Goodrich, "On using mixed-initiative control: a perspective for managing large-scale robotic teams," *4th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 165-172, 2009.

[21] Hesse, B.W., Werner, C.M., Altman, I., "Temporal aspects of computer-mediated communication," *Computers in Human Behavior* vol. 4(2), pp. 147-165, 1988.

[22] H. Sacks, E. Schegloff, and G. Jefferson, "A simplest systematic for the organization of turn-taking for conversation," *Language*, vol. 50, no. 4, pp. 696-735, 1974.

[23] M. L. McLaughlin, Conversation: How talk is organized, Sage Publications, 1984.

[24] J. Jaffe, and S. Feldstein, Rhythms of dialogue, Academic Press, New York, 1970.

[25] M. Yamamoto, and T. Watanabe, "Timing control effects of utterance to communicative actions on embodied interaction with a robot," *13th IEEE International Workshop on Robot and Human Communication*, pp. 467-472, 2004.

[26] B. Robins, K. Dautenhahn, R. Boekhorst,and C.L. Nehaniv, "Behaviour delay and robot expressiveness in child-robot interactions: a user study on interaction kinetics," *3rd ACM/IEEE International Conference on Human-Robot Interaction*, pp.17-24, 2008.

[27] Y. Okuno, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, "Providing Route Directions: Design of Robot's Utterance, Gesture, and Timing," *4th ACM/IEEE International Conference on Human-Robot Interaction*, pp.53-60, 2009.

[28] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "An affective guide robot in a shopping mall," *4th ACM/IEEE International Conference on Human-Robot Interaction*, pp.173-180, 2009.

[29] Stanney, K. *et al.*, "A paradigm shift in interactive computing: Deriving multimodal design principles from behavioral and neurological foundations," *International Journal of Human-Computer Interaction*, 17, pp. 229-257, 2004.

[30] K.R. Thórisson, "Natural turn-taking needs no manual," in *Multimodality in Language and Speech Systems*, pp. 173—207. Kluwer Academic Publishers, 2002.

[31] N. Mavridis, E. Machado *et al*.,"Real-time Teleoperation of an Industrial Robotic Arm Through Human Arm Movement Imitation", In *Proc. Intl. Symposium on Robotics and Intelligent Sensors*, Nagoya, Japan, 2010.

[32] D. Matsui, T. Minato, K. F. MacDorman, H. Ishiguro, "Generating Natural Motion in an Android by Mapping Human Motion?", *Humanoid Robots, Human-like Machines*, 2007.

[33] N. Mavridis, A. Tsamakos, N. Giakoumidis, H. Baloushi, S. Ashkari, M. Shamsi, A. Kaabi, "Steps towards Affordable Android Telepresence", in *Proceedings of the HRI 2011 Workshop on Social Robotic Telepresence*, 2011.

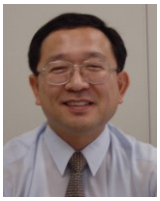
Dylan F. Glas received S.B. degrees in aerospace engineering and in earth, atmospheric, and planetary science from MIT in 1997, and he received his M.Eng. in aerospace engineering in 2000, also from MIT.

He has been a Researcher at the Intelligent Robotics and Communication Laboratories (IRC) at the Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan since 2005. From 1998-2000 he worked in the Tangible Media Group at the MIT Media Lab. His research interests include networked robot systems, teleoperation for social robots, human-machine interaction, ubiquitous sensing, and artificial intelligence.

Takayuki Kanda (M'04) received his B. Eng, M. Eng, and Ph. D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively. This author became a Member (M) of IEEE in 2004. From 2000 to 2003, he was an Intern Researcher at ATR Media Information Science Laboratories, and he is currently a Senior Researcher at ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. His current research interests include intelligent robotics, human-robot interaction, and vision-based mobile robots.

Hiroshi Ishiguro (M') received a B.Eng. and M.Eng. in computer science from Yamanashi University, Japan in 1986 and 1988, respectively. He received a D.Eng. in systems engineering from the Osaka University, Japan in 1991.

He is currently Professor in the Graduate School of Engineering at Osaka University (2002-). He is also Visiting Group Leader (2002-) of the Intelligent Robotics and Communication Laboratories at the Advanced Telecommunications Research Institute, where he previously worked as Visiting Researcher (1999-2002). He was previously Research Associate (1992-1994) in the Graduate School of Engineering Science at Osaka University and Associate Professor (1998-2000) in the Department of Social Informatics at Kyoto University. He was also Visiting Scholar (1998-1999) at the University of California, San Diego and Researcher at PREST of the Japan Science and Technology Corporation. In 2000 he founded Vstone Co. Ltd. He then became Associate Professor (2000-2001) and Professor (2001-2002) in the Department of Computer and Communication Sciences at Wakayama University. His research interests include distributed sensor systems, interactive robotics, and android science.



Norihiro Hagita (M'85 – SM'99) received his Ph.D. degree from Keio University (Japan) in 1986 in electrical engineering and joined Nippon Telegraph and Telephone Public Corporation (NTT) in 1978. He engaged specially in developing handwritten character recognition.

He also stayed as a visiting researcher at Prof. Stephen Palmer's lab in University of California, Berkeley (Dep. of Psychology) during 1989-1990. He is currently the director of ATR Intelligent Robotics and Communication Laboratories (IRC) at the Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan.