2013 IEEE RO-MAN: The 22nd IEEE International Symposium on
Robot and Human Interactive Communication
Gyeongju, Korea, August 26-29, 2013

Never Too Old for Teleoperation: Helping Elderly People Control a Conversational Service Robot

Dylan F. Glas, Kanae Wada, Masahiro Shiomi, Takayuki Kanda, *Member, IEEE*, Hiroshi Ishiguro, *Member, IEEE*, Norihiro Hagita, *Senior Member, IEEE*

Abstract— **The development of humanlike service robots which interact socially raises a new question: How can we make good interaction content for such robots? Domain experts specializing in the target service have the knowledge for making such content. Yet, while they can easily engage in good face-to-face interactions, we found it difficult for them to prepare conversational content for a robot in written form. Instead, we propose involving experts as teleoperators in an iterative development process in which the expert develops content, teleoperates a robot using that content, and then revises the content based on that interaction. We propose a software system and design guidelines to enable such an iterative design process. To validate these solutions, we conducted a comparison experiment in the field, with elderly volunteer guides teleoperating a robot at a tourist information center in Nara, Japan. The results showed that our system and guidelines enabled domain experts with no robotics background to create better interaction content and conduct better interactions than domain experts without our system.**

I. Introduction

Social robots, that is, humanoid robots which can provide services for people and interact with them in everyday social environments, have been a growing topic of interest for a variety of applications. Recent research has explored robots which can assist in nursing homes [1], provide route guidance in a shopping center [2], help people to do their shopping in a supermarket [3], greet people at a reception booth [4], and interact with people on city streets [5].

For many such applications, the complexity and open-endedness of conversational social interaction make full autonomy quite difficult to achieve. In many cases, researchers compensate for this by employing a human to teleoperate the robot for certain tasks, known as the “Wizard of Oz” method.

While this technique is often dismissed as a temporary measure for conducting experiments before certain technologies are available, we would like to propose a new possibility: that teleoperation of semi-autonomous robots presents an opportunity for elderly people to stay socially active and provide real value to society, even if they are bedridden or mobility-impaired.

Figure 1. Teleoperation of a robot interacting with a visitor at a tourist information center

A. Telepresence

Several kinds of telepresence technologies for social robots have been developed [6]. Many commercial robots such as the VGo telepresence robot show live video of the operator’s face, whereas others require the operation of controls or tracking of body motion to produce gestures and expressions through behaviors of the robot.

For our study, we chose a less direct method of teleoperation. Continually showing video of the operator’s face can be tiring for the operators, as they cannot relax, and it can be embarrassing if they are working at home and the video shows their personal space. Also, some people cannot speak clearly enough for a direct audio feed to the robot. For this reason, we chose to generate utterances based on textual content, through speech synthesis software, and to generate gestures using both implicit behaviors based on the textual content, and explicit behaviors specified by the operator.

Thus, the teleoperators in our study used a software interface to control a remote robot, as shown in Figure 1. Through this interface, content entered by the operator can be stored for reuse, so that the system becomes easier to use over time.

For future, autonomous systems, this kind of content generation through teleoperation may become a valuable way for experts to “teach” a robot how to conduct certain interactive tasks. So, while our current work focuses on employing elderly people as teleoperators for robots, it may be conceivable to employ them as robot trainers in the future.

* This work was supported in part by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research on Innovative Areas NO.21118001, and in part by the Ministry of Internal Affairs and Communications of Japan.

D. F. Glas, K. Wada, M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita are with the ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan 619-0288. D. F. Glas, K. Wada, and H. Ishiguro are also with the Intelligent Robotics Laboratory, Department of Systems Innovation, Osaka University, Toyonaka, Osaka, Japan, 560-8531. (Corresponding author: D. F. Glas. Phone: +81-774−95−2486, Fax: +81-774−95−1408; email: dylan@atr.jp).

B. *Communicative Knowledge*

In either teleoperation or training scenarios, the question arises: how can we develop good content (utterances and gestures) for such robots to provide information in a conversational style? It can be difficult for a programmer or researcher to make such content if they lack expertise in the target service domain. While people such as drama majors [4], could be considered experts in social interaction in general, they are not actually experts in a target application domain for the robot. It would be difficult to create content for a teacher, doctor, or salesperson robot, for example, without having the specific skills and knowledge for that job, even if one had access to factual information related to that domain.

This is because, aside from the **factual knowledge** required to perform a service effectively (e.g. knowing the route to the nearest convenience store), there is also **communicative knowledge** which is equally important. This is a special case of what is more generally referred to as “tacit knowledge” [7, 8]. For example, a teacher needs to know not only curriculum contents (factual knowledge) but also how to capture the attention of and motivate students (communicative knowledge). A doctor needs to understand the mechanisms of disease (factual) but also know how to talk to patients in a compassionate way (communicative).

While factual knowledge is often documented and relatively easy for a robot developer to acquire, communicative knowledge may be undocumented and can only be provided by a domain expert. In the field of Human-Robot Interaction, the concept of “design patterns for sociality” [9] encapsulates some common examples of everyday communicative knowledge that can be used in the design of social behaviors

C. *Teleoperation and Content Development*

One benefit of employing elderly teleoperators is that many of them have years of experience as a domain expert in some field. However, in early attempts to employ such experts make interaction content for humanoid robots, we discovered that it is not intuitive for domain experts to sit at computers and create textual content for a robot to provide a service in their own natural conversational style.

This is because communicative knowledge is often implicit and difficult to codify into explicit rules to store for future use. Thus, we believe that content incorporating this knowledge can be most easily created through an iterative process of content generation and conversational interaction through teleoperation, where the expert can use their communicative knowledge in an intuitive way.

We consider this teleoperation phase to be quite important in the content development process. As our previous studies revealed, situation coverage, representing the amount of knowledge stored in the robot, increases over time through teleoperation [10]. Content prepared in advance is typically premature, and through observing real people's reactions, an operator can identify content that is missing or needs to be changed. Hence, we believe domain experts can gain useful feedback through the process of teleoperation. We aim to integrate this step of real interaction through teleoperation into an iterative process for the development of interaction content.

D. *Scenario – Sightseeing Guide Robot*

In this study we consider the scenario of a robot providing information for tourists. The interaction content for such a robot would need to include a large amount of factual knowledge about the tourist attraction, as well as communicative knowledge like how to capture and sustain the interest of the tourists and tell stories in an engaging and exciting way.

The ideal domain experts who could provide this knowledge would be people currently working as guides at the target location in question. In Japan, there are many guide associations where senior citizens volunteer their time to work as guides and provide information to tourists about sightseeing attractions.

In this study, we worked with one such volunteer guide association in the city of Nara, Japan. The members of this group are all retired senior citizens, with an average age of 68.4. As walking around and providing information is physically demanding, they need to take time to rest and cannot work every day. Consequently they are often understaffed during busy seasons, and they were quite receptive to the idea of using robots to help reduce their workload.

Working with senior citizens provided some challenges, as most participants were not frequent computer users or fast typists. We did make accommodations for this, such as using large font and button sizes in our software interface. However, the focus of this study was on how to effectively enable them to apply their rich factual and communicative knowledge developed over years of experience as guides.

II. Related Work

A. *Dialog Construction*

In the studies of dialog, there have been a couple of dialog models developed. Typically, a dialog model assumes "tasks" in a dialog. That is, it assumes typical flow and/or set of information to be exchanged. For example, a dialog system for selling train tickets would expect information about departure time, destination, and number of passengers. For such task-oriented dialogs, a state-transition model or information-frame model fits well [11]. There are also authoring tools [12, 13], frameworks [14], and description languages [15, 16] that support the preparation of such task-based dialogs.

However, we aim to realize a chat-like conversation, where it is hard to anticipate a typical flow or set of information. There are agent-based models that can handle flexible dialogs, but it is difficult for people who are not experts in dialog systems to construct such dialog models [11]. Overall, these dialog studies did not reveal a way to convert knowledge from domain experts (who are not experts in dialog systems) into data useful for a robot's conversation.

TABLE I. MAIN GUIDELINES

Name	Main guidelines	For Design/Consolidation phase	For Operation phase
Responsiveness	React to what the visitor says during the interaction (make responses, change topics, greetings at the end)	(**A1**)Make behaviors* short	(**A4**) Use topic-independent utterances
		(**A2**)Write one idea in one behavior	
		(**A3**)Make topic-independent utterances	(**A5**) Watch and listen to the visitor carefully
Initiative	Lead the interaction (set expectations, initiate topics, minimize waiting time)	(**B1**) Design behaviors in a flow so the robot can lead the conversation	(**B2**) Choose behaviors smoothly at first
			(**B3**) Avoid typing too much
			(**B4**) Keep the conversation focused on prepared topics
Interactivity	Help the visitor participate in the interaction by asking questions	(**C1**) Make questions and prepare for the likely responses	(**C2**) Proactively ask questions

* Note: in this paper, we use the term "behavior" to refer to a combination of utterance and/or gesture. We primarily focus on utterances in this study, but at times we use the term "behavior," as our implemented system does support gestures as well as speech.

Toward the problem, an alternative approach would be the modeling of novice people's dialog. Chernova and her colleagues developed an on-line game to collect people's dialog, and converted the collected dialog data into a robot [17]. While such approach with a large dataset would be useful, it can often be difficult to collect data in advance, particularly if a dialog requires specific domain expertise that only experts would have.

B. Telepresence and Partial Autonomy

Previous studies have revealed a number of ways to provide service from distant locations. Telephone and video conferencing are widespread, and recently telepresence robots have also come into use [18, 19]. Studies have begun to investigate support techniques for telepresence robots [20]. In all of these telepresence approaches, it is a human user/operator who engages in a single channel of dialog.

In contrast, our approach uses partial autonomy, where multiple conversations can be supervised by a single operator [21]. In this approach, the ultimate goal is to let the system handle the majority of the dialog, with operators designated to support the system only when a situation is not covered by autonomy. While many situations can be automated for simple greeting and information-providing dialogs [22, 23], previous studies have not shown how to prepare and update dialog contents, or how to involve domain experts in the loop.

C. Guidelines for Dialogue Design

There are several research works which have focused on dialogue design for conversational agents. For example, Hollingsed et al. have investigated the effectiveness of the short-term response behaviors by using a tutorial system [24]. Moreover, Ward et al. have reported some usability issues in spoken dialog applications such as responsiveness, feedback and so on [25]. These research works have tried to identify important rules for dialog systems, but they did not focus on how to create a situation where domain experts can effectively create dialogue using their own intuitive rules.

By contrast, this paper aims to enable domain experts themselves to create content for a conversational robot.

III. INTERACTION GUIDELINES

In preliminary trials with the volunteer guides using early prototypes of our system, the iteration process was not as effective as we had hoped.

To help the operators, we analyzed the problems in their interactions and identified a number of common mistakes. We then developed a set of guidelines to assist the operators in content creation and teleoperation. These guidelines can be classified into three main categories as shown in Table 1: Responsiveness, Initiative, and Interactivity. Responsiveness is important at all times, whereas Initiative and Interactivity are complementary, and they must be balanced against each other.

A. Responsiveness

The first problem we observed was lack of responsiveness in the robot's interactions. One kind of poor responsiveness is when the robot responded to the visitor slowly, after a long silence. Sometimes this happened when the operator was taking time to search for a proper utterance from a list and didn't seem to feel any time pressure. Other times, the operator didn't find an appropriate utterance in the system and instead took a very long time to type a long utterance in response to a question.

Another problem with responsiveness is when the robot responded promptly, but appeared to ignore what the visitor was saying. For example, when the robot asked one visitor, "Where are you from?", and he answered, "I'm from Hokkaido!" the next utterance from a robot was "I'll explain about Nara." The visitor felt that the robot was not listening to what he said, or didn't care. Guidelines for responsiveness emphasize the importance of listening to the visitor and responding quickly and appropriately.

Reaction Time Studies have shown that delays longer than 2 seconds during interaction make people feel frustrated [26]. In teleoperation, an operator requires time to search through content or type utterances, so it can be difficult to respond so quickly. Our guidelines recommend that the operator react quickly to the visitor, and many features of our system (see Sec. IV) were developed to support this guideline.

Topic-independent utterances Many utterances can be classified under topics, such as "history of Nara Park." However, in natural conversation, people often use phrases such as, "oh, really?", "that's right", or "thank you," which do not fit into a topic. We use the term "topic-independent utterances" to refer to short utterances for making responses which cannot be classified within a topic. These include backchannel utterances such as, "uh-huh," "okay," "yeah, I see," as well as other miscellaneous phrases.

These utterances are usually necessary for a smooth dialog. Some behaviors serve to inform the speaker that the listener is paying attention and has understood what was said. They also play a role in turn-taking.

In our pre-trials, most participants prepared only factual, topic-specific utterances, but not topic-independent utterances. This resulted in awkward, one-way conversations where the robot seemed unresponsive to things the visitor said. Our guidelines explicitly recommend that operators prepare topic-independent utterances, as their use can lead to smoother, more natural interactions where the robot appears more responsive.

B. Initiative

In an ideal conversational situation, the dialogue literature would suggest that the robot and the visitor be given equal footing in terms of taking control of the conversation, and that a truly mixed-initiative system would result in better interactions than a fully robot-driven dialogue. However, there is an asymmetry in the system – for the operator to type a response to an unexpected question incurs a cost in terms of waiting time which would not occur in face-to-face conversation.

While a small number of such unprepared situations could be acceptable and informative, as they provide an opportunity for the operator to input new and useful content, too many unprepared situations will result in the robot's responses being unacceptably slow, as every response must be typed. Thus, operators need to take initiative, directing the conversation towards topics the robot can speak about. This should improve the robot's responsiveness.

Setting expectations In pre-trials, many participants designed behaviors only to say "Hello" or "Please ask me any question" at the beginning of an interaction, leaving the visitor confused as to what to do. As people interacting with the robot for the first time will not have clear expectations of the robot's abilities or role, our guidelines recommend that the operator start each interaction by establishing the robot's role and initiating the first topic of conversation.

Initiating topics Whenever the conversation stops progressing smoothly, the robot should initiate a new topic of conversation, rather than leaving this task to the visitor.

Minimize silence time If the robot makes the visitor wait too long, the visitor may choose not to wait for the robot to respond. To avoid uncomfortable silence, the visitor will often initiate an unrelated topic at such times. Just as in the "responsiveness" guidelines, the best way to avoid these situations is for the robot to respond quickly.

C. Interactivity

Many participants in our preliminary trials tended to create long monologues for the robot. Such one-sided interactions should be avoided, as visitors will feel bored and lose interest in the conversation. For example, a robot said "Let me tell you about Todaiji temple. Todaiji temple and the Great Buddha were first established in 743. This was because, due to earthquakes, hunger, and war, the emperor Shomu thought that Buddhism might be help to save the country. It needed too many …". This was too long of a one-sided explanation and boring for the people interacting with the robot.

Asking questions To enable an interactive conversation without exposing the robot to many unexpected questions, we recommend that the robot should actively ask questions to the visitor. This allows the visitor to participate in the conversation while the robot keeps the initiative. Also, replies to the questions are often predictable, so it is possible to prepare responses to expected replies.

For example, a robot explaining about the Great Buddha statue might ask, "How tall do you think the statue is?" instead of simply stating the statue's height. Visitors could respond with an estimate which is correct, too low, or too high, or they will simply say they don't know. Content can be prepared for each of these cases, e.g. if the guess is too high, the operator could prepare the phrase, "Well, it is very tall, but it's not THAT tall!"

IV. SYSTEM IMPLEMENTATION

We developed a software system to support operators in following the guidelines. In pre-trials, we observed that operators often did not follow the guidelines when we only told them orally. For example, some operators continued to type every behavior even after we told them "typing takes too much time, so please use the existing behaviors". Other operators continued to make long explanatory behaviors, even after we told them "please make behaviors short so you can see people's reactions and not make the interaction boring".

Graphical interfaces were designed for three distinct phases of use: the "design" phase, supporting content development; the "operation" phase, supporting real-time teleoperation; and the "consolidation" phase, supporting post-interaction review, including operator training and content improvement.

A. Design Phase

In this phase, operators prepare a basic set of content to enable simple conversations when operation begins. Until this basic content has been developed, it will not be possible to begin gathering feedback through teleoperation.

The primary tasks of the operator in the design phase are to create, edit, and organize interaction content for the robot. This content takes the form of "behaviors" which can include both utterances and gestures.

The design interface enables operators to enter utterance content, insert gestures if desired, and organize these behaviors into topics. An "add new behavior" wizard is provided which gives recommendations to the users, warning

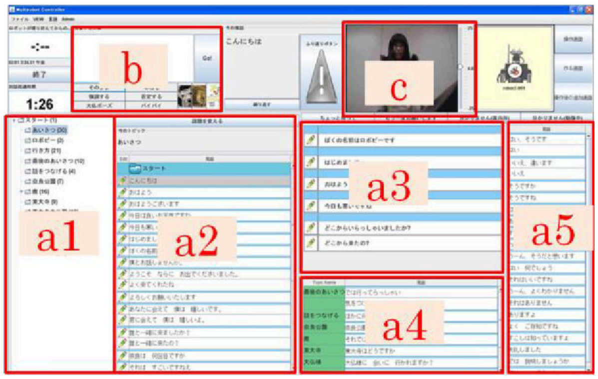

Figure 2. Operation view

them of behaviors which may be too long, or when there may be too many behaviors in one topic to allow easy searching.

B. Operation Phase

In this phase, operators teleoperate a robot using the content they have prepared. These interactions need to be conducted at a natural conversational speed for smoothness, and they should strike a careful balance between presenting prepared content and typing new content.

The Operation view, shown in Figure 2, enables the operator to teleoperate the robot in conversational interactions with people. Several features have been developed to support the operator in teleoperating the robot quickly and efficiently.

The basic interface contains a video feed from the robot (area "c" in Figure 2), a list of topics (a1), and a list of behaviors within the current selected topic (a2), which the operator can execute by clicking. Topic-independent utterances are shown in (a5) regardless of the current topic. The operator can also type text and choose gestures for new behaviors (b).

Shortcut features are provided to help reduce search time. Based on data from past interactions, the system predicts which behaviors are likely to come next and offers a list of best candidates (a3). It also shows links to behaviors which introduce new topics (a4), automatically switching the selected topic if one of these behaviors is chosen.

A "memo button" is provided for the operator's use in noting when a problem occurred during an interaction, marking a point in the interaction to be reviewed in the Consolidation phase. This button can be used when the operator would like to add further explanation, improve the phrasing of an utterance, improve the connection between explanations, prepare related behaviors, fix the robot's pronunciation, or change its gestures.

Finally, when the operator is slow in choosing or typing a behavior, the system helps to fill the delay by automatically inserting conversational fillers such as "hmm…" or "please wait a moment" to avoid awkward silences [26].

C. Consolidation Phase

In this phase, operators review the quality of the interactions and identify areas where the content base can be improved. This is achieved by reviewing a transcript and watching video of the interaction, and responding to system-generated prompts about recommended content changes.Immediately after an interaction is finished, the system presents a dialog asking the operator to self-evaluate their performance on a checklist of guidelines.

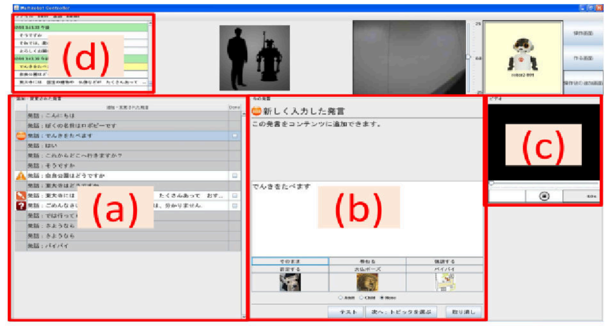

Figure 3. Consolidation view

The consolidation interface (Figure 3) guides the operator through the consolidation process. Recorded interactions are shown in (d), and a transcript of the selected interaction is shown in (a). In this list, entries representing "consolidation tasks" are highlighted, including behaviors created or edited and points where the memo button was pressed.

The central panel (b) acts as a wizard, guiding the operator through consolidation actions. For example, if a new utterance was typed, the operator can add it as a new behavior, edit it, or ignore it. If the memo button was pressed, the operator can add a new behavior or edit an existing behavior.

Finally, a video of the interaction can be viewed in (c). Operators tend to focus on operational tasks while controlling the robot, so reviewing a video of the interaction helps the operator to step back and watch the content and timing of the interaction itself. It has been shown that operators can have an impaired awareness of time during teleoperation [27].

V. EXPERIMENT

We conducted a field experiment to evaluate the effectiveness of the developed system and guidelines, in which we placed a robot in a tourist information center in Nara, with the task of explaining local sightseeing information to tourists.

A. Design

The experiment was designed to evaluate the hypothesis that operators with the proposed system and guidelines would make better content and operate the robot more effectively, resulting in a better overall impression, compared with operators who were not given our guidelines or the assistive software features.

To evaluate this hypothesis, we compared the performance of two groups of participants: one using our proposed system and guidelines, and one using a baseline system without our proposed features or guidelines. Participants operated the robot in interactions with real visitors at the tourist information center. We used a single-factor, between-participants experimental design, with conditions defined as follows:

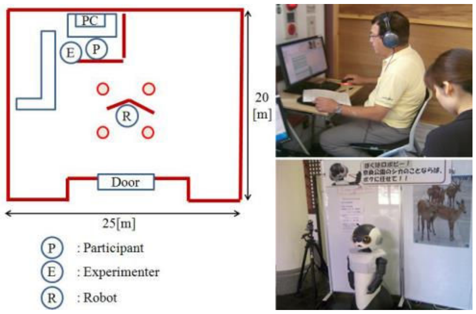

Figure 4. Field experiment environment

With-assistance: In this condition, we provided the guidelines and the developed system to participants.

Without-assistance: In this condition, we did not provide the guidelines or assistive features of the system designed to support the guidelines. Instead, we provided a basic system to allow them to enter content and operate the robot, and we allowed them to freely create and edit the content.

Specifically, the without-assistance condition did not include the "add new behavior" wizard, consolidation view, or guideline checklist. Of the features described in Section 4.2, links, topic shortcuts, the topic-independent utterance list, instances, and the memo button were not provided in the without-assistance condition. However, they were provided with a list of all behaviors entered during operation, and they were given the option to watch videos of their interactions. The interface of this simplified system is shown in Figure 5.

Automatic fillers were used in both conditions, and the experimenter answered questions from the participants about how to use the system in both conditions. For simplicity, gestures were not used in either condition.

B. Participants

A total of 27 participants (23 men and 4 women, who averaged 68.4 years old, s.d. 3.96) took part in our experiment. All were members of "Suzaku," a volunteer guide association in Nara. They each had 2-15 years of experience, and they were all currently active as volunteer guides at popular sightseeing areas at Nara at the time of the study. They had not previously interacted with our robot and had not had any experience operating any kind of conversational robots. Each participant provided their age and number of years of guiding experience, and we measured their computer ability, in terms of typing speed and speed of controlling a mouse. Based on this information, we assigned participants to conditions in order to balance these factors as closely as possible between conditions.

C. Equipment

In the experiment, we used the humanoid robot Robovie R3. It has a human-like appearance with two arms (2*4 DOF), a head (3 DOF), and is 110 cm tall. Its head has two eye cameras, a speaker, and a microphone. XIMERA software [28], was used for speech synthesis.

The operators in our experiments controlled the robot using the system described in Section 4, implemented in Java and running on a Windows PC. The interaction content created by the operator, including gestures and utterances, was stored in a database. When the operator chose a behavior for the robot to execute, the behavior contents were sent to the robot, which synthesized the utterance and executed the appropriate gestures.

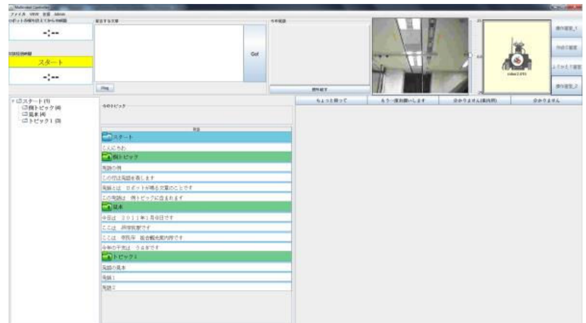

Figure 5. Graphical interface for Without-assistance condition

D. Procedure

In this experiment, the robot was to act as a guide specializing in talking about the deer in Nara Park, one of Nara's major sightseeing attractions. Its job was to talk with visitors and try to interest them in the deer, and to answer any questions they had about the deer. If visitors had off-topic questions, the robot was to direct them to the desk staff at the information center instead of trying to answer by itself.

The experiment consisted of two parts: preparation (the Design phase in our proposed flow) and operation (iterating through the Operation and Consolidation phases several times). One day was spent on each part. Thus, each participant took part in this experiment for two days.

The first day lasted for six hours at our laboratory: one hour of instruction on how to use the system, three hours to create interaction content, and two hours to practice by teleoperating the robot. On the second day, participants operated the robot for four hours at a tourist information center in Nara (Figure 4), presenting information and answering questions.

VI. Results

A. Interaction Quality

To measure the overall quality of interactions, two evaluators, blind to the experimental conditions, watched videos of the interactions and gave subjective quality ratings on a 100-point scale. This evaluation method was chosen instead of directly asking visitors for their impressions, due to the difficulty of getting consistent evaluations from first-time visitors; the robot is still novel and an interaction with the robot is still fun for many people, even with poor interaction content. Measuring the overall impression from a third-person perspective provides more consistent evaluations.

We asked evaluators and to rate its performance in its role as a guide providing information about deer in Nara Park, and how well it was able to engage in interactive conversation with the visitors. Scores were averaged over the final three interactions for each participant.

To provide a consistent scale for the evaluators, we gave reference definitions for 20-point increments, based on a scenario where the evaluator is an employer, choosing whether to hire the robot. In this scale, 100 is the best; 80

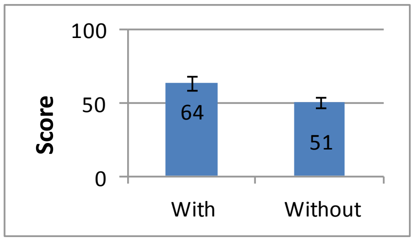

Figure 6. Overall evaluations of interaction quality

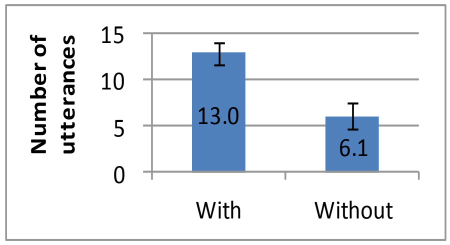

Figure 7. Number of utterances containing informational content

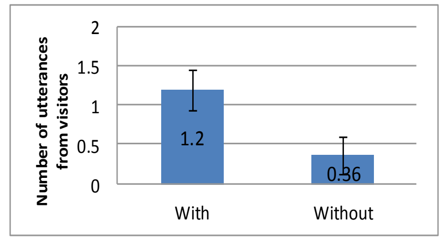

Figure 8. Number of visitor utterances showing surprise or interest

means that the evaluator feels that, as an employer, she could pay the robot slightly more than average; 60 is normal, and the evaluator feels that she could pay the robot slightly less than average; 40 is not good, and the evaluator feels she could only employ the robot without pay; 20 points is bad, but the robot could be forgiven by saying "I'm training"; and 0 points is unacceptable, where the evaluator felt she would not hire him even if he worked for free.

We computed the correlation of the overall impression scores between the two evaluators to be .645, which we consider to be a good match. Figure 6 shows the result of overall impression scores averaged between the two evaluators. A one-way factorial analysis of variance (ANOVA) revealed a significant main effect ($F(1, 25)=4.590$, $p=.042$, $\eta^2=.155$). Overall interaction quality was shown to be significantly better in the with-assistance condition, supporting our hypothesis.

B. *Secondary Measurements*

We counted the number of utterances containing informational content that were used by each operator. On average, 13.0 unique informational utterances were used per interaction in the with-assistance condition, compared with 6.1 in the without-assistance condition (Figure 7). A one-way factorial ANOVA revealed a significant difference between the two conditions ($F(1, 25)=13.278$, $p=.001$, $\eta^2=.347$))

We also counted the number of visitor utterances showing surprise or interest (e.g. "Wow!", or "Oh, really?") per interaction (Figure 8). An ANOVA revealed a significant main effect ($F(1, 25)= 5.69$, $p=.025$, $\eta^2=.185$). We interpret these results as indicating that operators in the with-assistance condition were able to conduct more interesting conversations.

VII. Discussion

A. *Contributions of Individual Components*

We introduced many techniques and guidelines at once in this study, and it is not clear to what extent each element contributed. The main focus of this study was to determine whether non-engineering domain experts could effectively be included in the content development process at all. A rigorous analysis of each system component is left for future work.

One major contribution of the software system appeared to be the video playback functionality. After watching videos of their operation, participants made a noticeably greater effort to operate the robot quickly and minimize visitor wait time. We attribute this to greater self-awareness, which enabled the operators to more effectively use their intuition and implicit communicative knowledge in interactions.

The "Topic-independent utterance list" was useful, enabling operators to react to the customers quickly in many situations. Other features, such as the automatic links and topic shortcuts, were not so important for small data sets like those used in our experiment, but we expect that their value will increase for larger sets of content and longer periods of operation (since links are built based on interaction history).

B. *Limitations*

This study only focused on making dialog. We did not consider locomotion or manipulation. The scale of this study was also relatively small, as each participant had only 3 hours for content creation and 3.5 hours for operation, covering only one topic. Long-term operation with a larger content set would make operation more difficult, but many of the proposed features in our system are designed to support large content sets.

C. *Applicability to Other Domains*

This study demonstrated that it is possible for nontechnical domain experts to create interaction content for a conversational robot through an iterative process of content development and teleoperation. Aside from guiding tourists, knowledge from domain experts might be necessary for robots working in a shop talking with customers, in a hospital or care home talking with patients, or in an educational setting helping students learn.

There will be application-specific differences. The communicative knowledge needed by a sightseeing guide robot centers around storytelling, engaging listeners, and reacting to their interests. Sales or education robots would have different strategies and goals. However, requirements such as smoothness of the interactions and responsiveness to the customer or student would be similar. Thus, we expect that our guidelines and system should be useful for such applications.

VIII. Conclusion

This paper addressed the challenge of employing elderly domain experts to teleoperate conversational service robots and develop interaction content. To this end, we proposed an iterative process using robot teleoperation in real interactions to provide feedback for improving conversational content.

We presented a system and a set of design guidelines to support domain experts in creating, using, and improving conversational content through teleoperation. We then evaluated how well a group of elderly volunteer guides could make conversational content and operate a robot using our proposed guidelines and system through a field experiment in a real tourist information center. The results confirmed that our system and guidelines helped operators conduct better interactions with the robot.

ACKNOWLEDGMENT

We wish to thank the staff at the Nara City Tourist information center, and the members of the NPO "Suzaku" for their helpful participation. We also wish to thank Satoshi Koizumi and Masaya Shimoyama for their help.

REFERENCES

[1] Pineau, J., Montemerlo, M., Pollack, M., Roy, N. and Thrun, S., "Towards robotic assistants in nursing homes: challenges and results", Robotics and Autonomous Systems, vol. 42, pp. 271-281, 2003.

[2] Shiomi, M., et.al., "Field Trial of Networked Social Robots in a Shopping Mall", IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2009), pp. 2846-2853, 2009.

[3] Gross, H.-M., et. al., "ShopBot: Progress in Developing an Interactive Mobile Shopping Assistant for Everyday Use", IEEE Int. Conf. on Systems, Man, and Cybernetics , pp. 3471-3478, 2008.

[4] Gockley, et. al.., "Designing Robots for Long-Term Social Interaction", *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (IROS2005), pp. 1338-1343, 2005.

[5] Weiss, A., Igelsböck, J., Tscheligi, M., Bauer, A., Kühnlenz, K., Wollherr, D. and Buss, M., "Robots Asking for Directions: The Willingness of Passers-by to Support Robots", ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2010), pp. 23-30, 2010.

[6] Kristoffersson, A., Coradeschi, S., and Loutfi, A., "A Review of Mobile Robotic Telepresence," *Advances in Human-Computer Interaction*, vol. 2013, p. 17, 2013.

[7] Polanyi, M., "Personal Knowledge: Towards a Post-Critical Philosophy". University of Chicago Press, Chicago, 1962.

[8] Eraut, M., "Non-formal learning and tacit knowledge in professional work", British Journal of Educational Psychology, 70:113-136, 2000.

[9] P. Kahn, N. Freier, T. Kanda, H. Ishiguro, J. Ruckert, R. Severson and S. Kane, "Design Patterns for Sociality in Human-Robot Interaction", *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2008)*, pp. 97-104, 2008.

[10] Kanda, T., et al., "A Communication Robot in a Shopping Mall", IEEE Transactions on Robotics, vol.26, pp.897-913, 2010.

[11] McTear, M. F., "Spoken Dialogue Technology: Enabling the Conversational User Interface", ACM Computing Surveys (CSUR), vol. 34, pp. 90-169, 2002.

[12] Glas, D. F., et. al, "An Interaction Design Framework for Social Robots", Robotics: Science and Systems Conference, 2011.

[13] McTear, M. F., "Modelling spoken dialogues with state transition diagrams:experiences with the CSLU toolkit",International Conference on Spoken Language Processing (ICSLP1998), pp. 1223-1226, 1998.

[14] Bohus, D., et. al., "Olympus: an open-source framework for conversational spoken language interface research," HLT-NAACL 2007 workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technology (2007), pp. 32-39, 2007.

[15] Krenn, B. and Sieber, G., "Functional Mark-up for Behaviour Planning: Theory and Practice", AAMAS Workshop FML: Functional Markup Language. Why Conversational Agents do what they do, 2008.

[16] Nishimura, Y., Minotsu, S., Dohi, H., Ishizuka, M., Nakano, M., Funakoshi, K., Takeuchi, J., Hasegawa, Y. and Tsujino, H., "A Markup Language for Describing Interactive Humanoid Robot Presentations", International Conference on Intelligent User Interfaces (IUI 2007), pp. 333-336, 2007.

[17] Chernova, S., Orkin, J. and Breazeal, C., "Crowdsourcing HRI through Online Multiplayer Games", the AAAI Fall Symposium "Dialog with Robots", pp. 14-19, 2010.

[18] Petelin, J. B., Nelson, M. E. and Goodman, J., "Deployment and early experience with remote-presence patient care in a community hospital", Surgical Endoscopy, vol. 21, pp. 53-56, 2007.

[19] Tsui, K. M., Desai, M., Yanco, H. A. and Uhlik, C., "Exploring Use Cases for Telepresence Robots", ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2011), 2011.

[20] Takayama, L., Marder-Eppstein, E., Harris, H. and Beer, J. M., "Assisted Driving of a Mobile Remote Presence System: System Design and Controlled User Evaluation", IEEE Int. Conf. on Robotics and Automation (ICRA2011), pp. 1883-1889, 2011.

[21] Glas, D. F., et. al, "Teleoperation of Multiple Social Robots", IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol. 42, No. 3, pp.530-544, 2012.

[22] Zheng, K., et. al., "How Many Social Robots Can One Operator Control?", in *ACM/IEEE 6th Annual Conference on Human-Robot Interaction* , ed. Lausanne, Switzerland., pp. 379-386, 2011.

[23] Kanda, T., et.al., "An Affective Guide Robot in a Shopping Mall", ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2009), pp. 173-180, 2009.

[24] Tasha K. Hollingsed and Nigel G. Ward, "A Combined Method for Discovering Short-Term Affect-Based Response Rules for Spoken Tutorial Dialog," ISCA ITRW Speech and Language Technology in Education(SLaTE), 2007.

[25] Ward, Nigel, Anais G. Rivera, Karen Ward, David G.Novick, "Some Usability Issues and Research Priorities in Spoken Dialog Applications", Technical Report UTEP-CS-05-23, 2005.

[26] Shiwa, T., et. al., "How Quickly Should Communication Robots Respond?", ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2008), pp. 153-160, 2008.

[27] Glas, D. F., et. al, Temporal Awareness in Teleoperation of Conversational Robots, IEEE Transactions on Systems, Man, and Cybernetics -- Part A: Systems and Humans, Vol. 42, No. 4, pp. 905-919, 2012.

[28] Kawai, H., Toda, T., Ni, J., Tsuzaki, M. and Tokuda, K., "XIMERA: A new TTS from ATR based on corpus-based technologies", ISCA Speech Synthesis Workshop, pp. 179-184, 2004.