A WOZ Environment for Studying Mutual Adaptive Behaviors in Gesture-based Human-robot Interaction

Yong Xu, Shinpei Takeda and Toyoaki Nishida

Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto
606-8501, Japan
{xuyong, takeda}@ii.i.kyoto-u.ac.jp, nishida@i.kyoto-u.ac.jp

Abstract

Mutual adaptation is considered to be one of the key issues in realizing a human robot interaction (HRI) system. In order to develop a mutually adaptive interactive robot system, the authors adopt a bootstrapping approach comprising three stages: a human-human WOZ (Wizard of Oz) experiment, a human-robot WOZ experiment, and a human-adaptive robot experiment. This paper will focus on the second stage where the authors developed a human-robot WOZ experimental environment and conducted an experiment to evaluate its effectiveness. The primary results suggested the following: the WOZ environment can function well; the operator’s view mode can be considered as a relatively intuitive operational mode for a human WOZ operator; and the realtime video image of the robot’s movement is necessary for the WOZ operator when he/she operates the robot while simultaneously watching the gestural instructions of the instructor. Some adaptive behaviors are observed as well. In addition, the primary results also suggested that the environment can be easily utilized with minor modifications to build a human-adaptive robot experiment.

Introduction

In recent years, the scope of robotics has extended well beyond industrial applications into several other fields including education, entertainment, and household appliances. In the near future, household robots and service robots are likely to become full-fledged family members. For example, as of May 2006, more than 2 million iRobot Roomba vacuuming robots have been sold worldwide (IROBOT 2007). The capacity of robot-environment-interaction is important for autonomous robots to accomplish simple tasks by interacting with the environment; however, the capacity of human-robot-interaction (HRI) for household or service robots such as nursing-care robots, guide robots, or waiter robots to perform cooperative tasks becomes more important. In order to create such robots, aspects such as “easy to use” or “easy to understand” must be emphasized so that non-specialists can easily understand and instruct such robots. Traditional human-computer interfaces may be inappropriate for this purpose since non-specialist users may experience difficulties in interacting with a robot using such interfaces. These interfaces may also cause unnecessary physical and/or mental stress in human users since they must adjust their behavior to the system. In order to realize a powerful interactive human-robot interface, the authors argue that mutual adaptation is a key feature to realize a user-friendly interactive human-robot interface. Mutual adaptation is a phenomenon where two interacting partners adapt to each other during their interaction. In the case of HRI, robots can typically adapt only to some specific people and learn their behaviors. In contrast, the authors attempt to realize a mutually adaptive robot that can not only adapt to any person but also utilize humans’ adaptability to dynamically establish communication protocols. In other words, the same instruction used by different people may contain different meanings, or different reactions are expected from the robot. In particular, when attempting to develop an interactive robot that can gradually adapt to human instructions through nonverbal communication, it is necessary to express the internal states of the robot such that mutual adaptation can occur.

This research adopts a bootstrapping approach that comprises three stages: a human-human WOZ (Wizard-of-Oz) experiment, a human-robot WOZ experiment, and a human-adaptive robot experiment. Instead of directly going to the third stage, we first make detailed observations of how people adapt to each other and how people improve the protocols for interacting with robots since it appears to be difficult to design a mutually adaptive robot system directly without any previous knowledge about mutually adaptive behavior. A WOZ experiment is a frequently used method in which participants interact with a system that is assumed to function autonomously. In fact, the system is usually fully or partially operated by a human operator. The goal of such experiments is to observe the behavior of
human users when they interact with the system rather than to measure the quality of the entire system.

The purpose of this paper is to outline the research strategy comprising three stages of mutual adaptation experimentation and show the effectiveness of our methodology by developing a human-robot WOZ experimental environment.

The remainder of this paper is organized as follows. Section 2 describes the related work on HRI. Section 3 overviews the three-stage strategy of this research. Section 4 describes the details about the experiment. Section 5 presents the conclusions and provides some directions for future work.

**Related work**

Jean Scholtz et al. (Scholtz 2003,2004) proposed a theory and developed a physical test arena for evaluating HRI; however, they focused on a rescue robot that requires greater mobility, robustness, and perception of a changing environment as opposed to the ability to establish communication with humans. Ulrich Nehmzow et al. (Nehmzow 2006) proposed a theory of robot-environment interaction and developed a robot system that can learn through task identification; however, they focus more on the autonomous function than on HRI. Bruemmer et al. (Bruemmer 2002) developed a game-style GUI (graphical user interface) to facilitate the interaction between a human operator and a robot. The potential applications of their research focused on various types of autonomous robots such as mine detector robots, rescue robots, and robots working in hazardous environment. In such applications, a human operator typically controls the robot remotely. They developed a classical user interface for different types of behavior-based robots and conducted some cooperative experiments. In these experiments, in order to improve the task performance, an initiative could be switched between a human operator and an autonomous robot so that the robot’s inferior capacity of obstacle avoidance and the human’s inferior ability of situation awareness could be complementary to each other. However, the authors prefer to develop a natural and intuitive human-robot interface rather than a traditional one.

This research will focus on studying a mutually adaptive human-robot interface system that can establish communication between humans and robots without predefined communication protocols. This feature will help humans feel the robots to be “easy to teach” by making them adaptive and being adaptable.

Several types of human-robot interactive systems have previously been developed. Hatakeyama (Hatakeyama 2004) developed a waiter robot system that can establish communication with human users by recognizing the intentions of some simple gestures according to a schema. Tajima (Tajima 2004) developed a software simulation system of a cleaner robot that can synchronize with the rhythm of a human user’s repetitive gesture. Ogasawara (Ogasawara 2005) developed a listener robot system that can listen to a human user’s explanation in a manner similar to what a human listener would do. All these robot systems can adapt to humans in specific tasks and situations but a general mechanism that can enable mutual adaptation between humans and robots has not yet been proposed. Although the embodiment (Ziemke 2001) of the robot may strongly affect the HRI results, the authors believe that if it is only necessary to verify the effectiveness of the HRI system itself, it may be sufficient to use a simple miniature robot that has a weaker embodiment rather than a humanoid robot, but still stronger than a software simulated robot.

**Three-stage Strategy**

In our research, a maze exploration task is considered as a common task for all experiments. There are two participants in the first two experiments and one in the last experiment.

In the first experiment, one participant (instructor) instructs another participant (actor) to move along a predefined path using only gestures. Although neither a robot nor a hidden operator exist in this experiment, it is still called a WOZ experiment since the actor plays the role of the robot by himself/herself.

In the second experiment, the instructor will instruct a robot to accomplish a similar task. The instructor is told that the robot can recognize his/her gesture. In fact, the robot will be controlled by another participant (operator). This is a genuine WOZ experiment where the WOZ operator plays the role of a gesture recognizer.

In the third experiment, the participant will instruct the robot directly, i.e., the robot has to recognize all the human instructor’s gestures and situations and perform all movements by itself.

**Human-human WOZ experiment**

In order to develop a mutually adaptive robot, it will be beneficial to first observe the mutually adaptive behaviors that occur in human-human communication. In the human-human WOZ experiment, each pair includes two participants (one instructor and one actor). Although there is neither a robot nor a wizard in the experiment, the actor can be considered to be a combination of a robot and a wizard.

The authors previously conducted a human-human WOZ experiment after some preliminary experiments (Xu 2007a). As a result, three findings, namely, alignment-based action, symbol-emergent learning, and environmental learning were obtained.

**Human-robot WOZ experiment**

An experimental environment was built for the second stage of our research. In this case, the participant (instructor) is asked to instruct the robot to move along a
predefined path that is projected on the ground. The instructor is told that the robot can recognize his/her gestures. In fact, another participant (operator) controls the robot while watching the instructor’s gestures through a monitor screen or through a magic mirror. The details about the environment are provided in the next section. Although only initial experiments were conducted, primary results can still partly prove the effectiveness of the WOZ experimental environment on the basis of several adaptive behaviors observed during the experiments.

**Human-adaptive robot experiment**

A human-adaptive robot experiment will be conducted in the third stage. A mutual adaptive algorithm is proposed by referring to the analysis results of the mutually adaptive behaviors observed in the previous two stages. In order to be human-friendly and easy to use, the human-robot interface should improve human users’ efficiency, relieve their psychological workload, and exclude their habitual behaviors, which often occur when they use standard interfaces. As a preliminary step toward a human-adaptive robot experiment, a one-way adaptive human-robot interface was developed as the first step (Xu 2007b).

**Experiment**

**Purpose**

The general purpose of the human-robot WOZ experiment is to develop an effective experimental environment so that the mutually adaptive behaviors occurring between the instructor and the operator can be observed. By analyzing the observed data, we hope to find some hints or clues for designing a general mutual adaptation algorithm. Furthermore, the difference between human-human gestural communication and human-robot gestural communication can be compared. In order to build an effective WOZ environment, we expect to achieve the following three sub-purposes:

- find an intuitive operation mode for the operator to control the robot
- verify if it is necessary for the operator to see the movement of the robot when he/she is controlling it
- verify if it is useful to enable the operator to see the robot’s movement without the map of the movement route.

**Task**

In this research, a maze exploration task is used for the experiment. There are two participants in the experiment; one participant (instructor) instructs a miniature robot (actor) to move along a predefined path, which is displayed on the ground, using only gestures. Another participant (operator) operates the mobile robot using joystick(s) while watching the instructor’s gestures. The instructor cannot see the operator and is told that he/she can instruct the robot directly through his/her gestures. After completing all tasks, all participants are asked to answer a questionnaire.

In order to achieve the three sub-purposes mentioned above, we defined three subtasks for different purposes.

- **Subtask A**: The operator was asked to control the robot without gestural instructions from the instructor to move along map 2, as shown in Fig. 1.
- **Subtask B**: The participants were asked to finish map 3, as shown in Fig. 1, and its upside-down version. In the first map, the operator can see the movement of the robot through a monitor using the images from camera 2 (Fig. 4), which made both the robot and the instructor visible to the operator. In the second map, the connection between the cameras and the monitor was changed so that the operator could only see the gestures of the instructor without the map on the floor along with the movements of the robot.
- **Subtask C**: The curve-type map shown in Fig. 1 was replaced by the matrix-type map shown in Fig. 2. Instead of being displayed on the floor, the map was printed on paper. The instructor could see the moving path on the printed map, but the operator could only see the number matrix without the moving path.

**Experimental environment and interface settings**

Since there are some differences between human-human interaction and human-robot interaction, the authors designed two different environments for human-human WOZ and human-robot WOZ, respectively. However, similar environments will be utilized for the human-robot WOZ experiment and the human-adaptive robot experiment.

In order to enable the WOZ operator to control the robot, an operator-robot interface is implemented with three operation modes: robot’s view mode, operator’s view mode, and tank mode.

In the robot’s view mode, only one joystick is used. The operator-robot interface comprises two joysticks (Saitek ST90). When the operator pulls the joystick along one direction, the robot will move along the same direction based on its own orientation.

In the operator’s view mode, the robot will move along the same direction as that of the operator instead of the robot. In the tank mode, two joysticks are used. Each wheel of the robot is controlled by one joystick. In this case, the left/right joystick control the left/right wheel of the robot. Pulling only one joystick forward/backward will cause the robot to turn left/right and pulling both sticks forward/backward at the same time can make the robot move forward/backward.

A projector (Epson EMP-740) is installed vertically on a rack which comprises two vertical poles and two horizontal poles. The projector is used to display the maps on the floor. Four cameras (Sony EVI-D30) are used to record the instructor’s gestures, the operator’s joystick operations,
and the movement of the robot from four vision angles. A CCD camera is fixed behind the projector for localization of the robot within the environment.

Two types of maps are used in this experiment: curve-type map and matrix-type map. Figure 1 shows four examples of the curve-type map and Fig. 2 shows one example of the matrix-type map.

The robot used in the experiment is a miniature robot (E-puck) designed by EPFL. Figure 3 shows the miniature robot (e-puck) with and without a localization marker attached from the top view and side view. Due to the low performance of the built-in CPU, the movement of the robot is controlled by a desktop computer running the WOZ software through a Bluetooth wireless connection.

The full configuration of the experimental environment is illustrated in Fig. 4. According to the operator’s subjective evaluation from the preliminary experiment, there was no obvious difference between the magic mirror and monitor. If the operator controls the robot from behind the magic mirror, it appears to be easily noticed by the instructor. In order to make the WOZ environment effective, only the monitor was used in all experiments.

Participants

Ten participants (three female and seven male) aged from 20 to 29 were asked to participate in the experiment. The average age was 23. For convenience, each participant was assigned and referred to using alphabet codes and numbers: A1, A2, B1, B2, C1, C2, D1, D2, E1, and E2. The participants were grouped into five pairs. Each pair was referred to using a group code (instructor code, operator code). Participants of the first three pairs A (A1,A2), B (B1,B2), and C (C1,C2) used the curve-type map to accomplish subtasks A and B. Pairs D (D1,D2) and E (E1,E2) used the matrix-type map to accomplish subtasks A and C. Pair D used a curve-type map without a cross symbol (Fig. 2), which indicated a bomb, and pair E (E1,E2) used a matrix-type map with a cross symbol.

Process

In subtask A, only one participant is asked to accomplish the task as an operator. The operator received the following simple instruction: “your task is to instruct the robot to move along the road shown on the floor by using gestures.” He/she was then given a simple sign to signify the start of the experiment.

In subtask B, the participant who played the role of instructor received the following simple instruction: “your task is to instruct the robot to move along the route printed on the paper by using gestures.” He/she was then given a simple sign to signify the start of the experiment.

The participant who played the role of operator received the following simple instruction: “your task is to control the robot’s movement; you can watch the instructor’s gestures through the monitor and control the robot using one or two of the joysticks according to your understanding of the instructor’s gestures.”

After finishing all subtasks, the participants were asked to answer a questionnaire about their subjective impressions of the robot, the WOZ environment, and the experiment.

In subtask C, the participant who played the role of instructor received the following simple instruction: “your task is to instruct the robot to move along the route printed on the paper by using gestures. You must avoid moving the robot over the cross symbol.” He/she was then given a simple sign to signify the start of the experiment.

In addition to the instructions received by the operator in subtask B, the operator was also told that: “you can press the button on the joystick to make the robot emit a sound to indicate that you cannot understand the meaning of the instructor’s gestures.”

After finishing all subtasks, all participants were asked to answer a questionnaire about their subjective impressions of the robot, the WOZ environment, and the experiment.

Results

The task completion time of the three subtasks were listed in Tables 1 through 3.
Figure 5 shows a typical scene of videotaped image data. The upper left shows the robot’s movement from the top view as recorded by camera 3; upper right shows the WOZ operator’s joystick operations as recorded by camera 4; and lower two images show the instructor’s gestures as recorded by cameras 2 and 1.

In subtask A, for finding an intuitive operation mode for the operator to control the robot alone, the average task completion time by all operators was shown in Table 1. Since the average completion time in the operator’s view mode (36.4 s) is significantly lesser than that in the other two modes, it suggests that the operator’s view mode is more intuitive than the other modes for the operators to operate the robot. However, the variances in the tank mode are significantly larger than those in the other two modes, which indicates that some people may prefer this mode more than the others. In fact, participant E said that he preferred the tank mode over the others, although his completion time using the tank mode was longer than that in the other modes.

In subtask B, for verifying the necessity of seeing the robot’s movement, the results of all pairs (Table 2) show that it was quite necessary for the operator to control the robot when he/she was watching the instructor’s gestures. It also suggests that the movement of the robot will provide the realtime feedback of joystick operation. The current state of the robot should be able to provide an important reference for both the operator and the instructor. The operator required the image of the robot so that he/she can know if his/her joystick operation reflects his/her intention of moving the robot, while the instructor will need to know if the robot correctly understands his/her meaning of previous gestural instructions and decide his/her next instruction according to the current situation by referring to the current state of the robot.

Subtask C was designed to verify if it is useful to enable the operator to see the robot’s movement without the moving path. Table 3 lists the results of completion times for this subtask. As shown in Fig. 6, gesture selection was observed by pair D (D1,D2). Instructor D1 did not initially know which types of gestures were available to instruct the robot, so she tried various types of gestures such as raising her arm and stepping. Since the WOZ operator D2 did not understand the meaning of such gestures, he pressed the joystick button to make the robot emit a sound. This informed the instructor that the robot could not understand her gestures; therefore, she tried several different gestures and found that a pointing gesture could be understood by the robot after receiving feedback about a failure on several occasions. From then on, she used only pointing gestures to instruct the robot.

From the results of the questionnaire, we obtained a subjective evaluation of the environment from the participants. Since no instructor participants realized that the robot was controlled by other people, it was inferred that the WOZ environment functioned quite successfully. We anticipated that the operator’s view mode may be the easiest mode of controlling the robot because the operator need not think about the mapping relation between the direction of the robot and himself/herself. The results for the task completion times of the operator’s view mode

Table 1: Task Completion Times for Subtask A (s)

<table>
<thead>
<tr>
<th>Operation mode</th>
<th>Average time (s)</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot’s view</td>
<td>42.4</td>
<td>72.3</td>
</tr>
<tr>
<td>Operator’s view</td>
<td>36.4</td>
<td>20.8</td>
</tr>
<tr>
<td>Tank mode</td>
<td>51.4</td>
<td>351.3</td>
</tr>
</tbody>
</table>

Table 2: Task Completion Times for Subtask B (s)

<table>
<thead>
<tr>
<th>Number of rounds</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pair A</td>
<td>227</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pair B</td>
<td>150</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pair C</td>
<td>106</td>
<td>51</td>
<td>36</td>
<td>38</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>Pair D (Can see robot)</td>
<td>94</td>
<td>69</td>
<td>48</td>
<td>42</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Pair E (Can’t see robot)</td>
<td>81</td>
<td>68</td>
<td>59</td>
<td>48</td>
<td>84</td>
<td></td>
</tr>
</tbody>
</table>

(“-” MEANS UNFINISHED IN TIME LIMITATION)

Table 3: Task Completion Times for Subtask C (s)

<table>
<thead>
<tr>
<th>Pair D</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Pair E</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round 1</td>
<td>144</td>
<td>98</td>
<td>102</td>
<td>Round 1</td>
<td>-</td>
<td>146</td>
<td>162</td>
</tr>
<tr>
<td>Round 2</td>
<td>82</td>
<td>91</td>
<td>63</td>
<td>Round 2</td>
<td>-</td>
<td>112</td>
<td>-</td>
</tr>
<tr>
<td>Round 3</td>
<td>68</td>
<td>66</td>
<td>58</td>
<td>Round 3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

(“-” MEANS UNFINISHED IN TIME LIMITATION)

Figure 3. The miniature robot (e-puck) used in the experiment

Figure 4. The configuration of human-robot WOZ experiment
exhibited the best values and thus proved the expectations. Figure 7 illustrated eight typical gestures among the twelve types used by instructor E1. Table 4 lists the number of times instructor E1 used gestures in subtask C. From the changing trend in the total number of gestures used, it can be found that the number of times a gesture was used by the instructor monotonically increased from 15 in the first minute to 21 in the fourth minute and decreased to 13 in the last minute. It suggested that the efficiency of the gestural communication was improved. Meanwhile, the types of gestures decreased from 12 at the beginning to 9 in the next minute and remained at 9 for the remaining duration. It might imply that after initial trial and error, the instructor and the operator might establish a relatively stable communication protocol so that the variety of gestures could attain a stable state. Table 5 shows the number of times and the rate of E1’s gestures. The maximum rate of hand shape or gesture action suggests that the finger pointing gesture is the type that is used most often. It indicates that the hand shape or action may also influence the effectiveness of gestural communication.

In subtasks B and C, the operator can make the robot emit a sound by pressing a button on the joystick. It can not only help the operator to express his/her understanding of the meaning of the instructor’s gestures but also help the instructor to know the internal states of the robot. Although only two pairs utilized this function, the questionnaire results suggested that it can potentially help both participants to establish communication between each other. Other features such as an LED flashing information may also possibly improve the capacity of expressing the internal states of the robot. Further experiments may be required in order to confirm the effectiveness of expressing such modality.
Part of the subjective feelings of the participants will be described here. Instructor E1 gave up complex gestures since he believed that the robot could not recognize it after experiencing the robot’s continuous sound (it implies that the robot cannot recognize the current gesture). This indicates that a miniature robot may not possess a sufficiently strong embodiment to convince a human of its cognitive ability.

With regard to the operator, some operators felt that it was difficult to simultaneously watch the gestures and the robot on different areas of the same screen. Operator E2 said he preferred the operator’s view mode to the tank mode, although he spent more time using the latter than the former. Considering his experiences of TV game using joysticks, it is inferred that the operator’s subjective evaluation of the traditional human-robot interface may affect his experience of using a similar interface.

### Conclusion and future work

In this paper, we introduced a human-robot WOZ experimental environment, which is the second stage of our three-stage bootstrapping approach, in order to develop a mutually adaptive interactive robot system. The approach comprised three stages: a human-human WOZ experiment, a human-robot WOZ experiment, and a human-adaptive robot experiment. The authors conducted an experiment to evaluate its effectiveness. The primary results suggested the following results: the WOZ environment can function well; the operator’s view is hopefully an intuitive operation mode for the WOZ operator; the real-time video image of the robot is important to the operator when he/she operates the robot while also watching the gestural instructions of the instructor. Some adaptive behaviors are observed as well. In addition, the environment can be shared easily to build a human-adaptive robot experiment. Since the number of participants is not sufficient to strongly support the authors’ hypotheses, we plan to conduct further experiments by adding more participants. The environment may also have to be improved so that more precise data can be obtained.

### References


