AI & Soc (2009) 23:201–212
DOI 10.1007/s00146-007-0134-1

ORIGINAL ARTICLE

WOZ experiments for understanding mutual adaptation

Yong Xu · Kazuhiro Ueda · Takanori Komatsu · Takeshi Okadome · Takashi Hattori · Yasuyuki Sumi · Toyoaki Nishida

Received: 15 August 2006 / Accepted: 10 April 2007 / Published online: 19 July 2007

Abstract A robot that is easy to teach not only has to be able to adapt to humans but also has to be easily adaptable to. In order to develop a robot with mutual adaptation ability, we believe that it will be beneficial to first observe the mutual adaptation behaviors that occur in human–human communication. In this paper, we propose a human–human WOZ (Wizard-of-Oz) experiment setting that can help us to observe and understand how the mutual adaptation procedure occurs between human beings in nonverbal communication. By analyzing the experimental results, we obtained three important findings: alignment-based action, symbol-emergent learning, and environmental learning.

Introduction

The recent "robot boom" has been caused by many well-known robots such as ASIMO (http://www.world.honda.com/ASIMO/) and AIBO (http://www.sony.jp/

Y. Xu (✉) · Y. Sumi · T. Nishida
Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
e-mail: xuyong@ii.ist.i.kyoto-u.ac.jp

K. Ueda
Department of General System Studies, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

T. Komatsu
Department of Media Architecture, Future University-Hakodate, 116-2 Kamedanakano, Hakodate, Hokkaido 041-8655, Japan

T. Okadome · T. Hattori
Innovative Communication Laboratory, NTT Communication Science Laboratories, 2-4 Hikaridai, Seika-cho, Keihanna Science City, Kyoto 619-0237, Japan

products/Consumer/aibo/) Human–robot interaction (HRI) will play an important role and attracts considerable attention in cases in which it is necessary for service robots to enter our daily lives and behave like members of our family. In order that a human user can teach a robot easily, it is not enough for the robot to adapt to humans one-sidedly; rather, the robot should also be able to help humans adapt to it by improving its capacity of adaptation and interaction. In other words, it is necessary to continuously build and develop a proper relationship between humans and robots by instilling in robots the capacity of mutual adaptation learning.

Several types of robots that are capable of cooperating or working along with people were developed (Nishida 2006) using the robots' *alignment ability*, through nonverbal interaction at various speeds. In our research, a robot with alignment ability implies that the robot can synchronize its behavior with instructions provided by a human user, e.g., when a user waves his/her hand quickly, the robot should also move quickly, and when the user slows down, the robot should do likewise. In the study conducted by Tajima (2004), a type of entrainment-based software robot was developed; this robot could react to human users' repetitive gestures by synchronization/modulation. In this case, *entrainment* is very similar to alignment. For example, an entrainment-based robot can not only synchronize with the orbit of the user's moving hand but also modulate to a new orbit after the user modifies the rhythm of his/her movement. Hatakeyama (2004) developed a schema-based robot that can appropriately respond to human behavior based on certain schemata. A *schema* is a kind of action sequence that is calculated on the basis of Bayesian networks. In Ogasawara (2005), a listener robot system was developed, which was capable of aligning with a human's actions based on nonverbal communication. The listener robot can establish natural joint attention with a human speaker using the redundancies of attention behaviors and acquire knowledge during the interactions. While these researches focused on improving the robots' adaptation abilities when interacting with humans, our research will focus more on how to facilitate humans' ease in teaching robots by improving the robots' mutually adaptive ability.

Through classical conditional learning, the researches of Yamada (2004) studied the way of helping an instructor search the conditions that AIBO was easy to learn. The learning in their experiment was a type of normal instruction-based learning. However, our research will be focused on mutual adaptation learning. When artifacts (e.g., robots) try to adapt to humans and learn their behaviors, humans may possibly change their behavior patterns simultaneously to adapt to the artifacts during the interaction between humans and artifacts. Meanwhile if the artifacts can improve their learning capacity gradually by taking usage of humans' high-level learning ability and expressing their internal learning status, it will make humans feel easy to teach or adapt to the artifacts. This type of capacity of artifacts is termed "mutual adaptation" in this paper.

Our final aim is to develop an interactive robot that can gradually adapt to humans' instructions using alignment-based learning through the nonverbal communication channel, and that is capable of mutual adaptation. However, it is difficult to develop such a robot with mutual adaptation ability without any knowledge of the mutual adaptation that is used in human–human communication.

Consequently, in order to develop a mutual adaptive human–robot interface, we designed an experimental environment to observe how and whether mutual adaptation occurs in human–human communication. We conducted an experiment and obtained three findings which were observed frequently: alignment-based action, symbol-emergent learning, and environmental learning.

Nonverbal communication-based mutual adaptive human–robot interface

Compared to industrial robots, it is more important for service robots to be able to interact with human users. Since verbal information processing causes a bottleneck with regard to the practicality of natural language processing technology, *nonverbal communication* is a good method for robots to achieve successful interaction with humans. On the one hand, it is natural for human users to use gestures to represent ambiguous intentions that are difficult to express verbally. On the other hand, it is technically easier for a robot to detect a user's hand movement than comprehend the exact meaning of a user's verbal message.

Mutual adaptation is an ability that makes a robot easily adaptable to a human user. Komatsu (2005) developed an experimental mutually adaptive speech interface in order to study mutual adaptation behaviors. This interface focuses on nonverbal communication, particularly that using phonological information. By sharing a common target, we prefer to extend Komatsu's research and focus on gestures. We chose a maze exploration task as a specific example in which an instructor and an actor cooperate with each other through nonverbal communication. In this task, the human instructor provides the actor (human or robot) with several directions to make the actor move according to a predefined map. The main aspect of this task is that the instructor can communicate with the actor only using hand gestures. Verbal language, facial expressions, and eye gaze information cannot be used. In this task, we believe that both the instructor and the actor should adapt to each other's actions (gestures or movements) so that they can reach the goal within the prescribed time limit. In order to develop a robot with mutual adaptation ability, we consider that it should be beneficial to providing theoretical bases through observing the mutual adaptation behaviors in the task.

Designing the experimental setting

Requirements for the mutual adaptation experiment

If mutual adaptation behaviors occur between the participants, these behaviors should be detectable from the time-series data that reflects the movements of both the participants. Although multiple communication channels may help participants to establish communication protocols with partners easily, this may make it difficult to distinguish the effect of each channel. In order to focus on a specific observable channel—hand gestures—we have to eliminate any possible interference from other communication channels, such as facial expressions, gaze direction, and voice.

Therefore, the following requirements are necessary for designing an experimental setting.

1. In order to obtain time series data, it is necessary to synchronize the data obtained from both the participants, i.e., the interaction partners.

In our experiment, synchronization was performed in a sensor room. Different kinds of data were synchronized—floor pressure sensor data, video camera data, USB camera data, and effect sound data.

2. In order to eliminate interference from other communication channels, it is necessary to limit the available channels between interaction partners to those that are observable.

In our experiment, we used sunglasses and masks to prevent interference from other nonverbal channels such as facial expressions and gaze direction. Since only gestures could be used, the actor had to move by watching the instructor's hand gestures.

3. It is necessary to monitor the learning process in order to observe alignment-based continuous value learning and symbol (protocol) emergent learning.

We designed a circulative route wherein the actors had to repeat the same route several times so that the learning process could be observed more clearly. In order to observe the continuous value learning, we designed a continuous route in an environment that had several objects in its path, including bombs, signals, targets, as well as the goal.

Research strategy for the mutual adaptation experiment

A WOZ (Wizard-of-Oz) experiment (Cheyer 1998) is a frequently used method. Our research strategy includes three steps: a human–human WOZ experiment, a human–robot WOZ experiment, and a human-learning robot experiment.

In order to develop a robot with mutual adaptation ability, we believe that it will be beneficial to first observe the mutual adaptation behaviors that occur in human–human communication. In the human–human WOZ experiment, each pair included 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. By analyzing the experimental results, we expect to obtain some useful knowledge that will be helpful in designing a learning robot as well as a human–robot interface.

Experiment

Purpose and setting

The purpose of the experiment is to observe mutual adaptation behavior. In order to meet the requirements—as mentioned in Sect. 3—we designed a human–human

WOZ experiment setting that could be used to observe the mutual adaptation behaviors of the participants. This setting is shown in Fig. 1.

With the actual position of the actor, the instructor would be able to perceive the entire task map with the positions of all the obstacles on the map, including bombs, signals, targets, as well as the goal and orbit. The instructor would also need to perceive the current, exact position of the actor. The reward, in the form of sound effects (clearing target objects, exploding bombs/signals, reaching the goal, etc.) and scores were necessary for both the instructor and the actor. The mask and sunglasses were used to prevent interference from other communication channels, such as facial expressions and gaze directions.

The graphical user interface (GUI) monitor software is shown in Fig. 2. The equipment used in this experiment is shown in Figs. 3 and 4.

The instructor observes the movement of the actor by viewing the floor sensor information displayed on the computer screen. In order to enable the instructor to observe the actor's movements more clearly, we installed a USB web camera. By comparing the image from the USB camera with the real-time-changed floor sensor information that reflected the actor's current footsteps, the instructor was able to judge whether the actor was moving along the correct orbit, stepping on a bomb, or approaching the target correctly. We also designed a signal object that appeared and disappeared once every 2 s in order to observe the changing rhythms of the instructor's gestures and the actor's movements.

Participants

Four male participants from 23 to 26 years of age participated in the experiment: one Chinese participant and three Japanese participants. All the participants were graduate school students. For convenience, each participant was assigned and referred to using alphabet codes: A, B, C, and D. We grouped the participants into

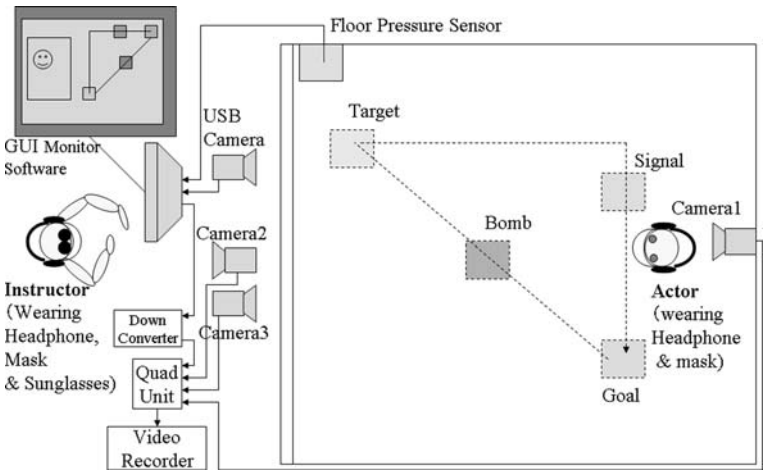


Fig. 1 Human–human WOZ experiment setting

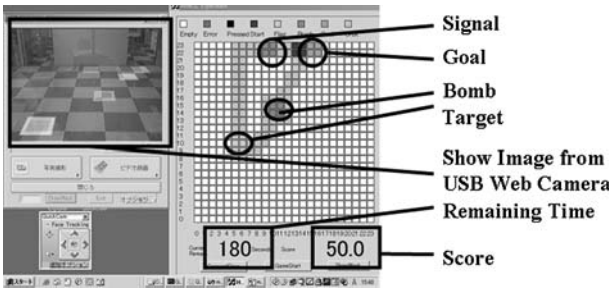


Fig. 2 GUI Monitor Software

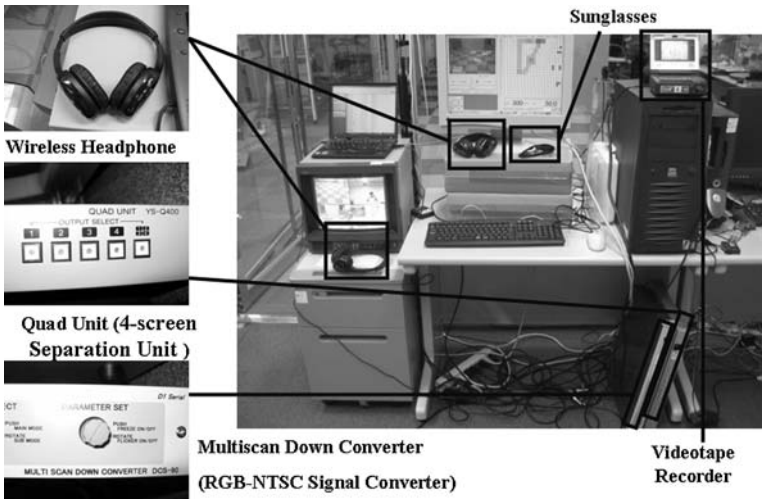


Fig. 3 Experimental equipment—1

six pairs and conducted experiments for 12 trials with two maps (for each pair, two maps were used for two trials) in the order of (A, B), (A, C), (D, C), (C, B), (D, B), and (B, D); each pair is expressed as (instructor, actor). In order to remove interference in establishing learning policy and communication protocol from sharing experience, the instructor and the actor were separated in different room during rest time so that conversations between them were prevented.

Procedure

The experimenter provided the instructor with the following set of instructions: "Your task is to instruct your partner (the actor) to move according to the map shown on the display screen by using only gestures. You should make your partner move along the orbit toward the flashing green target and try to reach the goal within the time limit, as quickly as possible." The experimenter gave the actor the following instructions: "Your task is to move by following your partner's

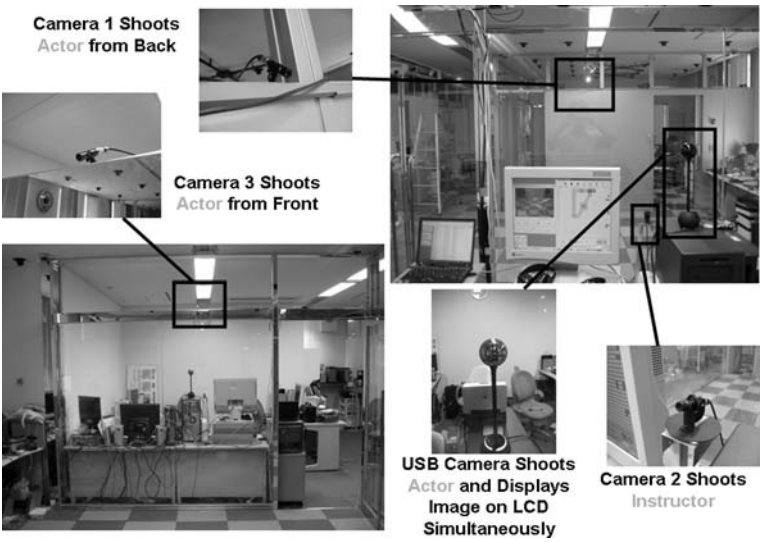
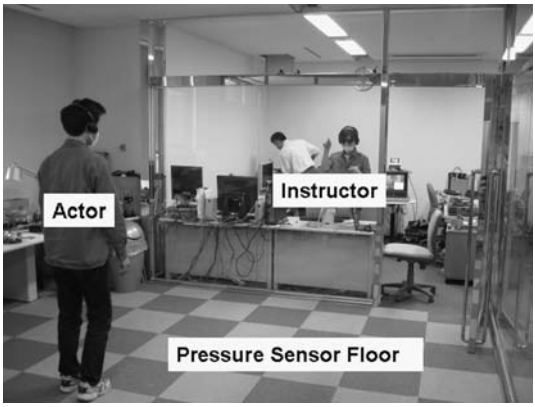


Fig. 4 Experimental equipment—2

(instructor) gestures and you should keep facing him for the duration of the entire experiment."

The experiment scene is shown in Fig. 5. When the experiment started, the software program played a WAV file with the following message (in Japanese): "3… 2… 1… start." If the actor moved to a position containing a bomb/an active signal—shown as a lit up/flashing pink box on the display—the sound of an explosion was played using the WAV file. Similarly, if the actor moved to a position containing the current target—shown as a flashing green box on the display—a ding-dong sound was played using the WAV file. Finally, if the actor moved to the position containing the goal within the stipulated time limit, the tune of "Congratulations!" was played. If the actor exceeded the time limit, the message

Fig. 5 Experiment scene



"Time over" was played. The time limit for every pair using each map was set to 5 min (300 s).

Results

After performing the human–human WOZ experiments, we obtained 12 video data files (with a total of 53 min of video data) and corresponding sensor log data files (including time stamp, floor sensor data, and effect sound). The video combined three images (camera 1, camera 2, and camera 3) with the output of the GUI monitor software. The GUI also included a real-time image captured by the USB camera. We analyzed the video data by annotating with a video annotation tool called ANVIL (Kipp 2004) and referring to the results of the sensor log data.

We found that the instructors preferred to use different behavior patterns when they instructed different actors, and there were similarities in the case of the actors.

Table 1 shows some of the results of the four pairs performing identical tasks using the same route and map. The rate of symbolic gestures (e.g., "stop," "keep going," "a bit") used by instructor A decreased from 39% in (A, B) to 26% in (A, C). Similarly, when D acted as an instructor, the rate changed from 50.4% in (D, B) to 23% in (D, C). Based on this, it can be inferred that the same instructor may have preferred to use different methods of instruction when dealing with different actors, i.e., there was probably an adaptation from instructors to actors. In addition, actor B moved in "short" steps most frequently—44.4%—in (A, B) and moved in "normal" steps most frequently—50.8%—in (D, B). Actor C took more "long" steps than "short" steps in (A, C) but took more "short" steps than "long" steps in (D, C). Since the correspondences between the instructor's gestures and the actor's movements are not pre-established, it can be inferred that the same actor may prefer to adopt different styles of movement when interacting with different instructors. This can be also considered as a case of adaptation from actors to instructors.

Table 1 Changing behavior patterns in crossing pairs

Instructor	Actor	Instructor's gesture type			Actor's step		
		Symbolic	Direction	Other	Long	Normal	Short
A	B	67	91	16	20	30	40
		39%	52%	9%	22.2%	33.3%	44.4%
A	C	33	87	8	28	49	9
		26%	68%	6%	33%	57%	10%
D	B	65	63	1	7	34	26
		50.4%	48.8	0.8%	10.4%	50.8%	38.8%
D	C	26	78	9	3	55	18
		23%	69%	8%	4%	72%	24%

Through a more detailed analysis of the video files, we obtained three important findings which were observed frequently: alignment-based action, symbol-emergent learning, and environmental learning.

Alignment-based action

Alignment-based action is a type of action in which two participants align their actions while interacting with each other. In our experiment, this implies that the actor changes his speed and/or steps in order to keep pace with the speed and/or width of the instructor's gesture, and vice versa. Based on the alignment action, two types of behaviors—pace keeping and timing matching—were observed, as shown in Fig. 6. Alignment action was observed to occur frequently (over 80%); thus, it can be considered as a basic communication protocol for all pairs.

Timing-matching behavior was observed in all pairs. In the first trial with pair (A, B), at the first occurrence of the passing signal, the instructor did not gesture until the signal disappeared, and the actor moved only after seeing the instructor's gesture. The time lag was approximately 0.4–1.0 s between the start of the instructor's gesture and the start of the actor's movement. Timing-matching behavior can be considered as another type of alignment-based action.

Symbol-emergent learning

The second type of mutual adaptive behavior is symbol-emergent learning, in which the instructor used gestures such as symbol-like instructions when he interacted with

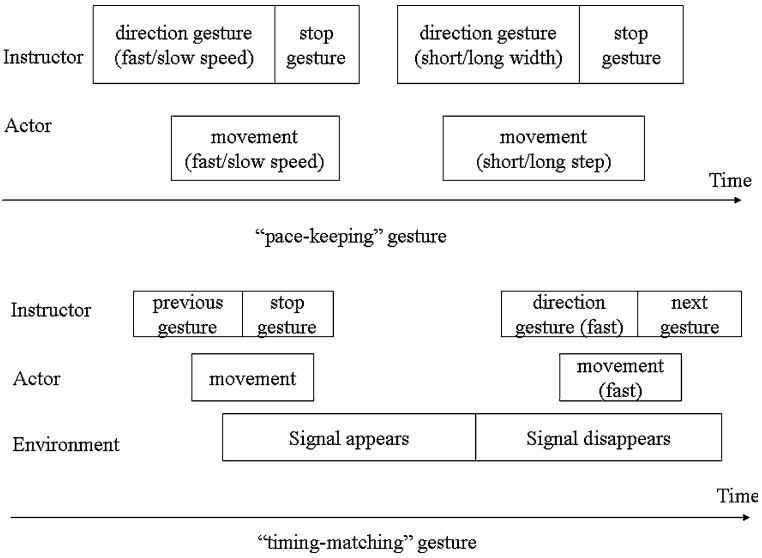


Fig. 6 "Pace-keeping" and "Timing-matching" gestures

the actor. The observed symbol-like instructions included the following gestures: "stop," "a bit," and "keep going."

The first type of symbol-like gesture—"stop"—comprised the following two types: when the instructor wanted to stop the movement of the actor, the instructor made a specific shape with his hand, as a result of which the actor stopped. The four participants used four different types of gestures for "stop" by making different shapes with their hands (including the "ok" sign, forward facing palm, downward facing palm, and fist). One of the characteristics shared by these "stop" gestures was that the instructors suddenly stopped their movement in the middle of making these gestures. It can be inferred that the movement of the hands may convey more meaning than the shape of hands. In other words, the alignment or synchronization may act as a basic principle when establishing the communication protocol. Another type of "stop" gesture was observed when the instructor stopped moving his hand and retained its shape and position; in this case also, the actor stopped and waited for a new gesture. A sudden change in the speed and/or shape of the hand gesture, however, might indicate "cancel" in addition to "stop," e.g., when an actor moved fast and returned to his previous position after seeing this kind of gesture. A slow change of gesture might indicate that the instructor wishes to switch to the next instruction.

The second type of symbol-like gesture—the "keep going" gesture—is observed in two situations, as shown in Fig. 7. In the first situation, the actor continued moving even after the instructor stopped moving his hand. This indicates that a pause in the gesture emerged as a symbol, possibly implying that the previous gesture was still applicable. In this case, establishment of a certain type of trusting relationship between the instructor and the actor may be a prerequisite. In the second situation, the actor continued moving because the instructor repeated the same hand gesture. Such a gesture could also be considered as a type of alignment-based action.

The third type of symbol-like gesture was that for "a bit." As shown in Fig. 8, when the instructor found that the actor was taking very long steps, he used the gesture for "a bit." After this, the actor increased the length of his steps at first, and when the instructor repeated the same gesture along with the direction gestures to signal to him to return to his previous position, he might have become aware that his previous understanding of "a bit" was wrong; then, he decreased the length of his steps to "normal" when he saw the gesture for "a bit" for the third time. Further, when the instructor used the gesture for "a bit" yet again, the actor moved in short steps. By using this type of symbol-like gesture, the instructor succeeded in changing the length of the actor's step as he desired.

Environmental learning

The last type of mutual adaptation behavior is referred to as "environmental learning." In the first trial of (A, B), the time spent by actor B on each round reduced as he moved five times along the same route. He spent approximately 7 s moving backward for approximately 3 m in the first round, and this decreased to 5 and 3 s in the third and fifth rounds, respectively, for the same distance. The most

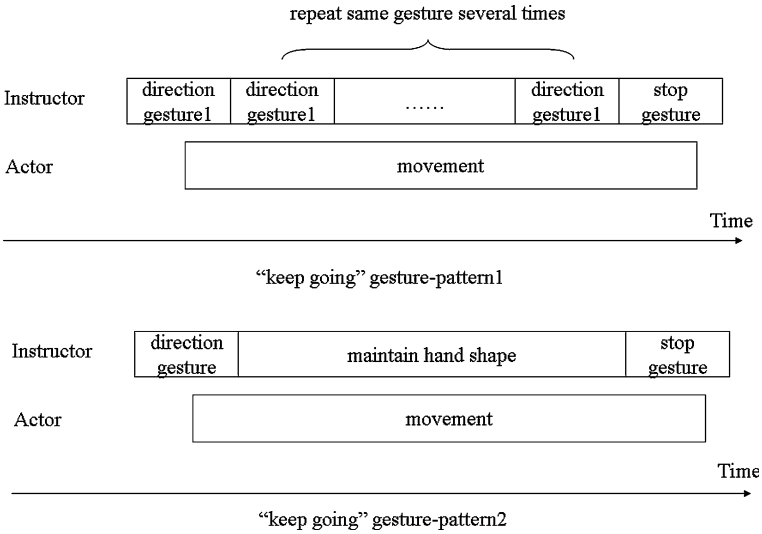


Fig. 7 "Keep Going" gesture

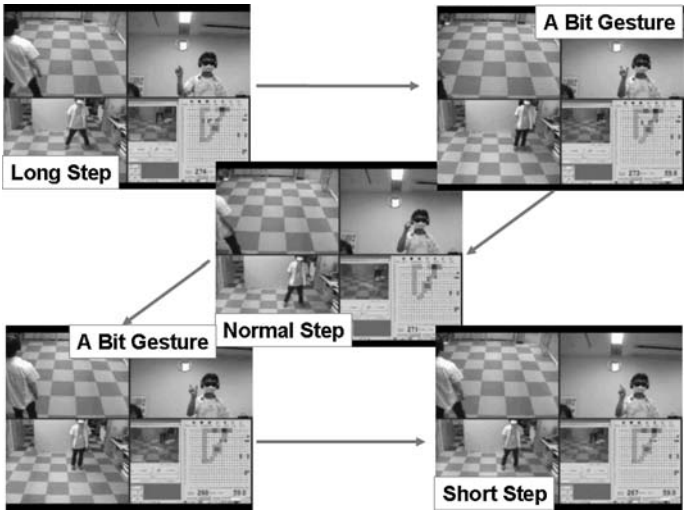


Fig. 8 "A Bit" gesture

plausible reason for this could be that the actor became familiar with the environment when he repeatedly moved along the same route; therefore, he moved with more confidence in the subsequent rounds. Furthermore, gestures of instructor A also decreased from 6 in first round to 3 and 2 in the third and fifth rounds, respectively. From this, it can be inferred that communication between the participants became smoother in the subsequent trials, and both participants adapted and improved their efficiency of movement.

Conclusion and discussions

In this paper, we discussed the method of designing a mutual adaptation experiment. By proposing an experiment setting, performing preliminary human–human WOZ experiments, and analyzing the results, we obtained some useful findings.

Although the number of participants was small, the findings are still informative and valuable. We consider that alignment-based action can be helpful to establish basic protocol in gesture-based human–robot communication. Symbol-emergent learning and environmental learning will help the robot to be easily adaptable to. Except for hand gestures, some head movements, such as head nodding or head shaking, were observed in few trials. These head movements may act as a positive or negative reward ("Yes" or "No") in some situations. We are going to conduct more experiments and give further analysis on these movements. We also expect to observe more typical mutual adaptation behaviors, including a set of multiple actions and adaptive selection of gestures by improving the setting and/or procedure of the experiment. In the human–robot experiment, we think it may be helpful to enable the robot's capacity to express its internal learning status by showing some meta-information in addition to showing by movement with different confidence.

With the exception of video data, we also plan to use motion sensors to record the instructors' hand gestures and head movements in order to develop gesture-measuring technology that will provide more useful information for data analysis. In addition, we intend to perform more experiments in the near future, including human–human and human–robot experiments.

References

Cheyer A, Julia L, Martin JC (1998) A unified framework for constructing multimodal experiments and applications. Lect Notes Comput Sci 2155:234–242

Hatakeyama M (2004) Human–robot interaction based on interaction schema (in Japanese). Master Thesis, University of Tokyo, Japan

Kipp M (2004) Gesture generation by imitation—from human behavior to computer character animation, Boca Raton, Florida: Dissertation.com

Komatsu T, Utsunomiya A, Suzuki K, Ueda K, Hiraki K, Oka N (2005) Experiments toward a mutual adaptive speech interface that adopts the cognitive features humans use for communication and induces and exploits users' adaptations. Int J Hum Comput Interact 18(3):243–268

Nishida T, Terada K, Tajima T, Hatakeyama M, Ogasawara Y, Sumi Y, Xu Y, Mohammad Y, Tarasenko K, Ohya T, Hiramatsu T (2006) Towards robots as an embodied knowledge medium, invited paper, special section on human communication II. IEICE Trans Inf Syst E89-D(6):1768–1780

Ogasawara Y, Okamoto M, Nakano IY, Xu Y, and Nishida T (2005) How to make robot a robust and interactive communicator. In: Khosla R et al (eds) KES, LNAI, vol 3683, Part III3683, pp 289–295

Tajima T, Xu Y, and Nishida T (2004) Entrainment based human–agent interaction. In: Proceedings of 2004 IEEE conference on robotics, automation and mechatronics (RAM2004), Singapore, pp 1042–1047

Yamada S, Yamaguchi T (2004) Training AIBO like a dog. In: The 13th international workshop on robot and human interactive communication (ROMAN-2004), Kurashiki, Japan, pp 431–436