The Sixteenth International Symposium on Artificial Life and Robotics 2011 (AROB 16th '11),
B-Con Plaza, Beppu,Oita, Japan, January 27-29, 2011

User study of a life-supporting humanoid directed in a multimodal language

T. Oka[1], T. Abe[2], K. Sugita[3], and M. Yokota[3]

[1] *Nihon University, 1-2-1 Izumicho, Narashino, Chiba, 275-8575 JAPAN*
[2]*Nihon Computer Kaihatsu Co. Ltd., 6-24-9, Minami-Ooi, Shinagawai-ku, Tokyo, 140-0013, JAPAN*
[2]*Fukuoka Institute of Technology, 3-30-1, Wajiro-higashi, Higashi-ku, Fukuoka, 811-0295, JAPAN*
Tel : 81-47-474-9693; Fax : 81-47-474-2669
oka.tetsushi@nihon-u.ac.jp, t-abe@nck-tky.co.jp, {sugita, yokota}@fit.ac.jp

Abstract: This paper describes a user study of a life-supporting humanoid directed in a multimodal language and discusses the results. Twenty inexperienced users commanded the humanoid in a computer simulated remote home environment in the language by pressing keypad buttons and speaking to the robot. The results show that they comprehended the language well and were able to give commands successfully. They often chose a button press action in place of verbal phrases to specify a direction, speed, length, angle, and/or temperature value, and preferred multimodal commands to spoken commands. However, they did not think that it was very easy to give commands in the language. This paper discusses the results and points out both strong and weak points of the language and robots.

Keywords: Life-supporting robot, multimodal language, user study, humanoid, human-robot interaction

I. INTRODUCTION

It is predicted that in near future life supporting robots will help people in homes, streets, hospitals, offices, etc. In order to design and develop such robots, one should take into account both cost and user friendliness, since such robots must be affordable for people in need and designed for those untrained.

Although menu-based conventional GUIs are cost effective, they are not suited for untrained people and through such interfaces one cannot communicate with robots in a natural manner even if they look like humans. It is understandable that we expect robots to communicate with us just like ourselves. In fact, there are thousands of robots which can speak and understand verbal messages [1]. However, verbal communication is just one aspect of human communication; we use gestures, postures, eye contacts, paralanguage, etc. to convey different kinds of information [2]. Therefore, we can well image that humanoids will communicate with us both verbally and nonverbally. Obviously, such robots will be more user friendly than robots operated through a conventional user interface [3]. Unfortunately, in order to communicate like humans, robots will need many kinds of sophisticated sensors to perceive verbal and nonverbal signals, high-performance computers to disambiguate messages and infer users' intensions [4], and an articulated body which look and move smoothly like a human. For this reason, precise imitation of a human will prevent us from developing affordable user friendly life-supporting robots.

The authors have designed RUNA (Robot Users' Natural Command Language), a multimodal command language [5-8], taking into account both cost effectiveness and natural human-robot communication. In RUNA, one can command an action without ambiguity by specifying its type and parameter values. Several versions of RUNA, which combine spoken verbal messages and nonverbal messages such as hand gestures, body touch actions, and button press actions have been developed and evaluated with novice users.

A small humanoid that can be directed in an earlier version of RUNA was investigated [5]. Novice users remotely commanded the humanoid in RUNA to explore a room. All the users completed their task with help from a three-page leaflet and most of them conceived a fairly high opinion of the robot and the language, although there were some shortcomings and limitations in the robot. The overall command success rate during the task was about 70 % and the robot reacted to noises due to our poor cheap microphone. Because of the limitation of the onboard computer, the robot had to choose among 80 actions of walking, turning, looking around, and so on; it were able to turn by 30 degrees but not by 40 degrees, so it rejected some grammatical commands in RUNA. It was difficult for novice users to inform the robot one out of five turning angle values by pressing a button for a definite period of time; they had no idea how long they should press it to turn the robot as much as they want, because they were given no instructions or opportunities to practice.

After the study, we developed a simulated humanoid with a more sophisticated command interpreter and eliminated some of the shortcomings found in our first humanoid. The new robot has a larger repertoire of actions, including 540 left turns. We redesigned a set of button press actions to specify action parameter values such as speed, angle, length, temperature values.

This paper presents a user study of the new life-supporting humanoid. All of the twenty users completed their tasks, to explore a remote room and operate an air conditioner, successfully commanding the robot in RUNA. The success rate of each user's commands was over 90 % and several of the problems in the old humanoid were resolved. Most of the users preferred multimodal commands to single modal spoken

commands and often selected a button press action to specify parameter values. However, the users did not think that it was very easy to command the robot in the new version of RUNA. The results shed light on some problems of our multimodal language as well as advantages. Most importantly, it will be easier for novice users to communicate with a life-supporting robot, if they can convey parameter values one by one in nonverbal messages and can always omit some parameter values which are obvious in the context or values which they do not care about. Since there are cases in which it is hard to determine parameter values beforehand, users should be allowed to modify commands at any time.

II. MULTI-MODAL LANGUAGE

In the study presented in this paper, we used a version of the multi-modal language, RUNA [6], which comprises a set of grammar rules and a lexicon for spoken commands, and a set of non-verbal events detected using keypad buttons. The spoken language enables users to command a humanoid in Japanese utterances, completely specifying an action to be executed. Commands in the spoken language can be modified by nonverbal events.

An action command in RUNA consists of an action *type* such as *walk*, *turn*, *report*, and *lowertemp* (for lowering the temperature setting) and action *parameters* such as *speed*, *direction*, *angle*, *object* and *temperature*. Table 1 shows examples of action types and commands. The action types are categorized into 24 classes based on the way action parameters are specified in Japanese.

There are more than 300 generative rules for the full version of RUNA. These rules allow Japanese speakers to command robots actions by speech alone. In RUNA, a spoken action command is an imperative utterance including a verb to determine the action type and other words to specify action parameters. There are more than 250 words, categorized into about 100 groups identified by non-terminal symbols.

Table 1 Examples of action commands

Type	Command	English Utterance
walk	*walk_s_3steps*	Take 3 steps slowly!
turn	*turn_f_l_30deg*	Turn 30 deg. left quickly!
move	*move_m_r_2steps*	Move 2 steps right!
look	*look_f_l*	Look left quickly!
raisetemp	*raisetemp_room_2deg*	Raise the temperature of the room by 2 degrees!
settemp	*settemp_aircon_22deg*	Set the air-conditioner temperature around 22 degrees!
query	*query_aircon_all*	Report the status of the air-conditioner!

In RUNA, non-verbal events modify the meaning of spoken commands. They convey information about parameters of action commands. Table 2 shows examples of non-verbal events; users can use keypad buttons to specify action parameters values instead of mentioning them. This reduces average number of words in a command and speech recognition errors. One can command a robot saying "turn" and pressing a button simultaneously instead of saying "turn 60 degrees left slowly!" Furthermore, multimodal commands are often more natural than spoken commands: e. g. pointing a glass and saying "pick this up" or saying "lower the temperature" pressing a button.

If a button event has been arrived within a short period of time, a spoken command will be modified as shown in Table 2. The twelve buttons are assigned to specific parameter values (Fig. 1). The direction and speed of a turning action command are determined by the key pressed most recently by the user. A single key press action conveys an angle value (10, 20, 40, 50 or 60 degrees) and a step value (1, 2, 4, 5, or 8 step(s)) depending on the duration, while a multiple key press action conveys values (90 or 180 degrees and 10, 20, and 30 steps). Likewise, the robot will make the preset temperature *two* degrees higher, if a key has been pressed *twice* before a spoken command "raise the room temperature!"

Finally, the repeat button and query button allow users to command robots without speaking. The empty button convey default parameter values, such as 3 steps, 30 degrees, right, normal speed, etc.

Table 2 Button event and action parameters

action type	duration	count	button
sidestep/walk etc.	distance	distance	speed direction
turn etc.	angle	angle	speed direction
look etc.	-	-	speed/target
raise/lowertemp	-	temperature	-

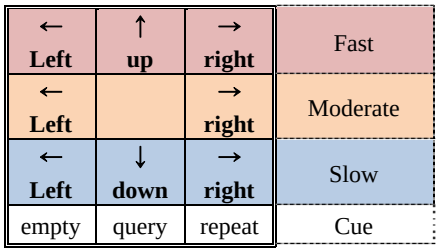

Fig. 1 Key assignment for action parameters

III. USER STUDY METHOD

We conducted a user study of a humanoid which can be commanded in RUNA using a remote interface. Five users who had experienced another version of RUNA and 15 novice users (male and female, aged 13-30 years) commanded the robot to achieve two different tasks: exploring a remote room and operating an air conditioner in the room. Before commanding the robot, the users watched a short demonstration movie for 70 seconds and read an eight-page document illustrating how to give commands in RUNA in diagrams and figures for five minutes. Then, we explained them how to give spoken and multimodal commands showing the

same document. They were allowed to practice commanding the robot for up to 20 minutes.

In order to test their comprehension and competence, we also gave them a comprehension test and asked them to give some extra spoken and multimodal commands precisely as printed in sheets of paper. In addition, they were asked to answer questions about the robot and our multimodal language at the final stage.

IV. RESULTS

Table 3 summarizes the data of each user. Users 1-15 commanded the robot in RUNA for the first time, while users A-E had already contributed to our previous studies. All the users completed their tasks: they answered three questions about the remote room correctly and switched on, changed the temperature setting, and switch off the air conditioner as instructed.

As shown in Tables 4 and 5, they gave many spoken and multimodal commands to move forward and turn the robot to explore the virtual remote home. They often repeated to move the robot forward or turn it to the same direction as if to look how the camera view image on the screen changed. Some users gave repetition commands to do so by pressing the "repeat" button in Fig. 1, but the other users did not. For some types of actions there were more multimodal commands than speech only commands (Table 5). Table 6 shows how the users specified angles and distance values. About 70% of the commands were given within ten seconds after the previous command was completed.

Table 3 Representative data of each user

ID	SR	nc	nw	time	test	Q1	Q2	Q3	Q4
A	100	42	1.0	7:30	8	6	7	M	Y
B	100	48	2.3	9:55	8	4	3	M	N
C	97	33	3.4	11:16	8	4	4	M	Y
D	100	21	2.4	6:07	7	2	3	S	Y
E	100	41	2.5	10:00	10	4	3	M	Y
1	100	36	1.4	8:20	8	6	4	M	N
2	100	31	2.0	12:00	8	4	6	M	Y
3	100	23	2.1	10:37	8	2	3	M	Y
4	100	33	2.0	19:50	10	5	5	M	Y
5	100	38	1.9	9:45	10	5	5	M	Y
6	100	43	1.5	12:10	9	5	4	S	Y
7	100	27	2.1	8:05	10	4	5	M	Y
8	100	26	2.9	8:03	10	4	4	S	Y
9	99	74	1.9	23:06	8	6	7	M	Y
10	98	43	1.8	10:43	9	5	5	M	Y
11	97	31	3.8	11:35	6	6	6	M	Y
12	96	31	4.2	9:20	7	4	4	S	Y
13	96	25	3.7	11:10	7	2	2	S	Y
14	91	23	2.2	13:30	6	4	5	M	Y
15	90	31	2.5	11:46	8	3	5	S	Y
ave	-	35	-	11:20	8.3	4.3	4.5	-	-

SR: Command success rate nc: Number of given commands
nw: average number of words time: task completion time
test: comprehension test result before the tasks (ten questions)
Q1. Did you command the robot in a natural way? (7 pt. scale)
Q2. Was it easy to command the robot? (7 pt. scale)
Q3. Which do you prefer, spoken or multimodal commands?
Q4. Is this robot helpful for you? (Do you want it?)

At the beginning, some of the users consulted the eight-page document, but they did it less frequently at the end. Some users seldom turned the pages or took time to look at the diagrams.

Table 6 shows that the users were poor at conveying one of five values by the duration of a button press action even after achieving their tasks, while it was straightforward for them to give parameter values by pressing a button twice or three times. The users selected *quick/fast* actions most frequently and there were more multimodal commands to turn the robot to the right than to the left for some reasons (Table 7).

After some practice, most of the users spoke clearly and fluently in most cases. However, they failed to convey action types and/or parameter values by speech for several reasons. Some users hesitated to give commands including many words a few times. Seven users failed to change the temperature setting of the air conditioner using a wordy speech only command, saying "change the temperature setting of the air conditioner to 23 degrees" and failed instead of giving a simpler multimodal command, pressing a button twice and saying "lower the room temperature!" There were also ambiguous and *ungrammatical* commands such as a button press action followed by an utterance "lower it!"

A few users tried to modify or restate a command and failed while the robot was executing an action. Some users used words which are not included in the lexicon of RUNA. Some spoken messages were utterly clear and fluent but misrecognized by the speech recognition system. Finally, Table 8 shows some important comments from users.

Table 4 Modality choice for parameter values

type	value	duration	count	speech	default
turn	0-30 deg.	85	-	8	49
	31-60 deg.	17	-	18	-
	61-180 deg.	-	31	29	-
walk & move forward	0-9 steps	30	-	36	16
	10-50 steps	-	109	34	-
	0-99 cm	-	-	1	-
	1-3 m	-	-	5	-
	"much"	-	-	1	-

Table 5 Number of commands given by users

action type	spoken	multimodal	button	total
walk/forward	56	174	-	230
backward	2	1	-	3
turn	92	147	-	239
sidestep	18	24	-	42
look	12	9	-	21
look around	10	0	-	10
switch on/off	45	-	-	45
query	15	-	11	26
settemp	18	-	-	18
lowertemp	1	4	-	5
repetition	-	-	53	53

Table 6 Success rates of parameter specification

modality	Value	Success rate (%)
duration	0-150[ms]	76.2
	700-1300[ms]	81.0
	1300-2000[ms]	47.6
count	2	100
	4	100

Table 7 Parameter specifications using a button

parameter	value	specifications
speed	fast	251
	moderate	81
	slow	9
direction	left	38
	right	75

Table 8 Users' comments

C1	It was difficult to specify a parameter value by the duration of a button press action.
C2	It was hard to measure the distance from the robot to a target in the camera image.
C3	It was difficult to specify an angle value verbally or nonverbally.
C4	It was difficult to command the robot to move forward and turn.
C5	It was difficult to change the temperature setting.
C6	I wish if I had more time to practice.
C7	I could not speak to the robot fluently.

V. DISCUSSION

The results show that the users understood the language well and mostly succeeded in giving commands during the tasks (Tables 3 and 5). They often chose a button press action and specified direction, speed, and temperature values without difficulty (Tables 5 and 7). They preferred to use buttons and thought that the robot was helpful (Q3 and Q4 in Table 3). Each user spoke only one to four words and well avoided speech recognition errors, slips of the tongue and wordy commands. We presume that it gets easier for users to specify action parameter values using a button and they will choose button press actions more often, since cognitively speaking it is easier to press a button than to generate verbal phrases.

It is obvious that they did not think that the language was a very natural or easy one to communicate with the robot (Q1 and Q2). Surprisingly, this result is worth than the previous study [5]. Although it is difficult to find the good reasons in the data, we can point out some disadvantages of the language which might have caused difficulties for the beginners. First, there are action parameter values, such as angles and walking steps, which are difficult to determine before giving a command (C2 in Table 8); many users chose short turns using a single press and 10-30 step walks using a multiple press possibly because they did not know the right values. They did not know how to turn the robot to a target or how much to move the robot forward to reach a target point. Therefore, robots should allow users to modify parameter values at any time. It would be better if one can send a cue to stop the robot at the right place. Second, they were unable to choose among angle or step values using a single press action (Table 6 and C1 and C3), none the less because they were given some time to practice. Although users may adapt, there should be ways around for beginners. Third, we suspect that it is difficult for beginners to specify two or more values in a single button press action, because in the previous study button press actions specified one or two values while in this study users had to specify two or three values at once without omission. Therefore, it would be better if one can specify values one by one and leave out parameters which are not important. The language will be more natural if robots can infer users intentions based on the context. In fact, novice users often left out words whenever it was natural in daily communication. In addition, beginners may hesitate or restate commands, so life supporting robots must be able to deal with hesitant or halting spoken commands.

We think that the current version of RUNA is too complicated for novice users and has some defects which make the language a little awkward. The language should cover as many natural verbal commands as possible. We also suspect that the eight page document might have been a little confusing. We should not force users to learn what words and phrases they can say.

The results of our studies show that multimodal commands are advantageous in some ways and preferred by novice users of life supporting robots. We have realized a cost effective humanoid novice users can successfully direct to a position and make operate an air conditioner. The robot will be more user-friendly if one can communicate with it in a simpler, more natural, and flexible language.

AKNOWLEDGMENT

This work was supported by KAKENHI Grant-in-Aid for Scientific Research (C) (19500171).

REFERENCES

[1] Prasad R, Saruwatari H, Shikano K (2004) Robots that can hear, understand and talk. Advanced Robotics 18-5:533-564

[2] Knapp ML, Hall JA (2010) Nonverbal Communication in Human Interaction. Wadsworth

[3] Perzanowski D, et. al. (2001) Building a multimodal human-robot interface. IEEE Intelligent Systems, 16-1, pp. 16-21

[4] Jurafsky D, Martin JH (2000) Speech and Language Processing. Prentice Hall

[5] Oka T, Abe T, Shimoji M, Nakamura T, Sugita K, Yokota M (2008) Directing humanoids in a multi-modal command language. The 17th International Symposium on Robot and Human Interactive Communication

[6] Oka T, Abe T, Sugita K, Yokota M (2009) RUNA: a multi-modal command language for home robot users. Journal on Artificial Life and Robotics 13-2: 455-459

[7] Oka T, Abe T, Sugita K, Yokota M (2009) Success rates in a multimodal command language for home robot users. Journal on Artificial Life and Robotics 14-2:219-223

[8] Oka T, Sugita K, Yokota M (2010) Commanding a humanoid to move objects in a multimodal language. Journal on Artificial Life and Robotics 15-1:17-20