RUNA: A multi-modal command language for home robot users

T. Oka, T. Abe, K. Sugita and M. Yokota

Fukuoka Institute of Technology, Fukuoka JAPAN
(Tel : 81-92-606-4813; Fax : 81-92-606--0574)
(oka@fit.ac.jp)

Abstract: This paper describes a multi-modal command language for home robot users and a robot system which interprets users' messages in the language through microphones, visual and tactile sensors, and control buttons. The command language comprises a set of grammar rules, a lexicon, and non-verbal events detected in hand gestures, readings of tactile sensors attached on robots and buttons on controllers in users' hands. Prototype humanoid systems which can execute commands in the language are also presented along with preliminary experiments of human-robot interaction and teleoperation. Subjects unfamiliar with the language were able to command humanoids after a brief demonstration or with a two-page document at hand.

Keywords: command language, multi-modal, home robots, speech, gestures, tactile sensing

I. INTRODUCTION

Home robots that help ordinary people are expected to understand what their users want them to do as soon as commands are given. For people who need helps, even computer GUIs or remote controls of TVs are not ideal interfaces. Some may give up using the robots before learning which commands all the buttons, sliders and levers are linked with. In addition, if the robots are to execute many kinds of commands, their users will have to learn long sequences of operations.

A spoken language interface can be a good interface as it is not necessary for users to learn a new language. However, difficult problems of natural language understanding, speech processing, etc. must be solved before realizing such robots in practical use [1, 2].

Recently, a number of robot systems implementing automatic speech recognition have been realized [2]. To reduce computational cost, many robot systems interpret what the user means by means of keyword spotting, creating semantic representations of user utterances. In those systems, it is not clear to users what kind of an utterance is understood or not since they are not based on a well-defined language. It is also difficult to identify the meaning by keyword spotting alone if a variety of commands are given in different situations.

It is believed that non-verbal interfaces have advantages especially in communication of quantitative and spatial information and help natural interaction between humans and robots. Home robots will have many sensors and actuators over their body which can be thought of as devices for non-verbal communication. However, multi-modal communication integrating verbal and non-verbal communication between humans and robots has not yet been fully explored by researchers in the related fields.

For the above reasons, the authors proposed to design and use a practical multi-modal language to command home robots, integrating a simple but well-defined spoken language and non-verbal messages [3].

This paper presents the first version of a multi-modal command language, RUNA (Robot Users' Natural Command Language), which enables users to command home robots speaking to them using gestures and touching in a natural way. A robot system which can execute commands in the language and preliminary experiments with small humanoids are also illustrated. Subjects unfamiliar with the language were able to command humanoids when they were given a brief demonstration or a two-page document

II. RUNA, the multi-modal language

The multi-modal language comprises a set of grammar rules and a lexicon for spoken commands, and a set of non-verbal events, i.e. events detected using visual and tactile sensors on robots and events from controllers in users' hands. The spoken language itself enables users to command home robots in Japanese utterances, completely specifying an action to be executed. Spoken commands can be modified by non-verbal messages including gestures, touching and button pressing.

1. Commands and actions

In RUNA, an action command consists of an action type such as *walk, turn, and move* and action parameters such as speed, direction and angle. Table 1 shows examples of action types and commands in RUNA.

The 23 action types of RUNA are categorized into twelve classes based on the way action parameters are specified in Japanese (Table 2). In other words, actions of different classes are commanded with different modifiers in RUNA.

Table 1 Action Representation of RUNA

Type	Command	English Utterance
walk	*walk_s_3steps*	Walk 3 steps slowly!
turn	*turn_f_l_30deg*	Turn 30° left quickly!
punch	*punch_s_lh_str*	Punch straight with the left hand!
wave	*wave_s_lh*	Wave your left hand slowly!
hug	*hug_s*	Give me a hug slowly!

Table 2 Action Classes and Types in RUNA

Class	Type	Parameters
1	*stand*	*speed*
	lie	
	squat	
	crouch	
	hug	
2	*moveforward*	*speed, distance*
	movebackward	
3	*walk*	*speed, steps*
4	*look*	*speed,* *target*
	turnto	
	lookaround	
5	*turn*	*speed, directionlrni, angle*
6	*sidestep*	*speed, directionlrni, distance*
7	*move*	*speed, directionni, distance*
8	*handshake*	*speed, handed*
	highfive	
9	*punch*	*speed, handed, directionni*
10	*kich*	*speed, footde, directionni*
11	*turnbp*	*speed ,* *bodypartwo,* *directionni*
	mavebp	
	raisebp	
	lowerbp	
12	*dropbp*	*bodypartwo*

2. Syntax of RUNA

Table 3 shows some of the 171 generative rules for spoken commands in RUNA. These rules allow Japanese speakers to command robots actions in a natural way by speech alone, even though there are no recursive rules. In RUNA, a spoken action command is an imperative utterance in Japanese including a word to determine the action type and other words to specify action parameters. For instance, a spoken command "Yukkuri 2 metoru aruke! (Walk 2m slowly!)" indicates an action type *walk* and the distance *2m* (also see Fig. 1 for the parse tree). The third rule in Table 3 generates an action command of class 2 (AC2) which has *speed* and *distance* (SD) as parameters.

There are 166 words, each of which has different pronunciation, categorized into 60 groups (or parts of speech) identified by non-terminal symbols (Table 4). Note that users can use different words for the same action type or parameter value. Because the language is simple, well-defined and based on Japanese, Japanese speakers do not need any training to learn it.

In RUNA, one commands robots in shorter utterances touching them, using gestures or pressing buttons on controllers, specifying actions in more natural ways. A set of non-verbal events are defined for multi-modal action commands. Table 5 lists up representative non-verbal events in RUNA, which can be distinguished by their type, direction, speed, length, body part, etc.

2. Semantics of RUNA

In order to execute commands in RUNA, intermediate representation of actions shown in Table 1 is employed. Thanks to the simplicity of the spoken language, it is straightforward to identify the action type and parameters in a spoken command if correctly recognized by a speech recognizer.

RUNA allows users to omit action parameters although the action type cannot be left out when a new action is commanded. If some of the parameters of the specified type are missing in a spoken command, default values are assigned to those parameters. For example, "Kick!" is interpreted as "Kick straight with your right foot slowly!" in our command language (Table 6).

Non-verbal events play an important role in the semantics of the multi-modal language. They modify the meaning of temporally close spoken commands by replacing unspecified default parameter values.

Table 3 Grammar rules of RUNA

No	Rule	Description
1	S → Action	action command
2	S → Modifier	action modifier
3	Action → SD AC2	class 2 command
4	AC2 → AT_WALK	action type *walk*
5	SD → SPEED	speed
6	SD → DIST SPEED	distance + speed
7	SD → DIST	distance
8	Modifier → REPEAT	repeat last action

Table 4 Part of RUNA's Lexicon

Non-terminal	Terminal	Pronunciation
AT_WALK	at_walk_aruke at_walk_hokou at_walk_hokoushiro	aruke hoko: hoko:shiro
REPEAT	md_repeat_moikkai	mo:iqkai
SPEED	sp_fast_hayaku sp_fast_isoide sp_slowly_yukkuri	hayaku isoide yuqkuri
LUNIT	lu_cm_cm lu_m_m lu_mm_mm	seNchi me:toru miri
DIR_LR	dir_right_migi dir_left_hidari dir_left_hidarigawa	migi hidari hidarigawa
NI	joshi_ni_ni	ni
DEICTICBP	bp_deictic_koko	koko

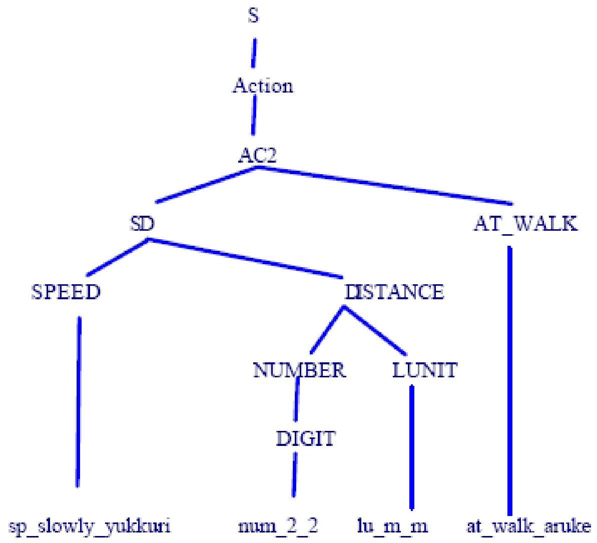

Fig.1. Parse tree for "Walk 2m slowly!"

Table 5 Non-verbal events of RUNA

Type	Event	Description
gesture	hm_left_slow_long	hand motion
gesture	hm_right_fast_short	hand motion
tactile	tc_rightwrist_long	touching
tactile	tc_leftshoulder_short	touching
button	btn_left_long	button press

Table 6 Default action parameter values

Type	Speed	Dir	Ang	Dist	BP
walk	slow	NA	NA	3steps	NA
look	slow	straight	NA	NA	NA
turnto	slow	right	90	NA	NA
turn	slow	right	30	NA	NA
wave	slow	straight	NA	NA	r-hand

III. Prototype Robot System

A prototype system based on Open Agent Architecture (OAA, http://www.ai.sri.com/~oaa/) [4] has been developed (Fig.2) and tested in some preliminary experiments, which is running on a desktop PC with a 2.4 GHz Core2Duo processor and 2 GB memory. It comprises eight OAA agents written in Java and C++. Julian, (a grammar-based large vocabulary continuous speech recognition engine http://julius.sourceforge.jp/), is employed for understanding of spoken commands in the ASR agent.

The event detectors monitor sensor readings and send non-verbal events to the interpreter agent. To detect hand motion gestures, an agent is implemented using Intel's Open Source Computer Vision Library (http://www.intel.com/), which detects the direction, speed, amplitude and frequency of hand waving motions. Tactile and control button events are also detected by other non-verbal event detectors.

The interpreter agent determines the action type and parameter values of each command based on speech and non-verbal events. It looks for recent non-verbal events and modifies a spoken command whenever it arrives from the ASR agent. An action database is consulted for default parameter values (as seen in Table 6).

The motion selector agent looks into a motion database to select robot motion patterns that match action command strings created by the interpreter agent. In the prototype system, 70 motion patterns are registered in the database, each of which can match more than one action string: e.g. both move_slowly_forward_3steps and walk_slowly_3steps match the same motion pattern.

IV. Applications

Two small humanoids (Kyosho Manoi PF-01) were used in some preliminary experiments with subjects unfamiliar with the command language.

One of the humanoids has a wireless camera and can be operated by a users monitoring what it is seeing on a remote PC, currently by speech alone. It was tested with 25 subjects including high school and university students who had never operated the robot. The subjects were able to remotely direct the robot to find a basketball in an unknown environment.

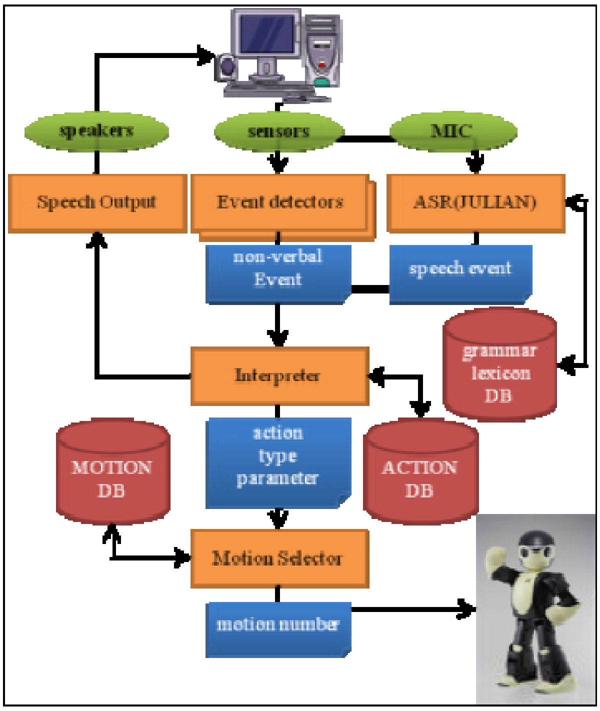

Fig.2. OAA-based prototype of RUNA System

The other humanoid has eight tactile sensors on its shoulders, arms and wrists, a microphone and camera attached on a PC. It was tested with 20 subjects including elementary school pupils. After a brief demonstration, they were able to make the robot, punch, kick, walk, turn, shake hands using multi-modal commands.

In the experiments, the speech recognizer performed better when the grammar was reduced for the specific tasks removing particular words and rules. Tactile events helped the system correctly identify the direction of commanded action as shorter utterances are easier to recognize for the speech recognizer.

V. Discussion

Our command language is under development for a more generic command language for multi-purpose home robots as discussed in the previous work [3]. Home robots in future will be given a diversity of tasks including taking care of home devices, internet access, and physical tasks. The current version is context free except for repetition of the last action commanded, and this reduces the computational cost of language understanding.

Although the system's usability has not been fully studied, the authors expect that non-verbal modalities and simpler spoken messages will help effective human-robot communication if the language is adapted to potential users. It is necessary to find effective non-verbal events and map them to action parameters so that users can specify them without mentioning them.

VI. Summary

A multi-modal language based on a simple spoken language and non-verbal event detection was designed to command home robots. A robot system which interprets and executes multi-modal commands in the language was developed and tested. Subjects without knowledge about the language or robot system were able to direct humanoids.

AKNOWLEDGMENT

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

REFERENCES

[1] Bos J & Oka T (2007), A spoken language interface with a mobile robot. Journal of Artificial Life and Robotics, 11-1:42-47

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

[3] Oka T & Yokota M (2007), Designing a multi-modal language for directing multipurpose home robots. Proc. of the 12th International Symposium on Artificial Life and Robotics (AROB 12th '07), Beppu, Japan

[4] Cheyer A & Martin D (2001), The open agent architecture. Journal of Autonomous Agents and Multi-Agent Systems, 4-1/2:143-148, March