The Fourteenth International Symposium on Artificial Life and Robotics 2009 (AROB 14th '09),
B-Con Plaza, Beppu, Oita, Japan, February 5 - 7, 2009

Voice-based Control of a Robotic Forceps by Using Displayed Image and Auxiliary Information

Keigo Watanabe, Takuya Tokunaga, and Kiyotaka Izumi
Department of Advanced Systems Control Engineering
Graduate School of Science and Engineering, Saga University
1-Honjomachi, Saga 840-8502, Japan

Abstract

A method of using an image related to three-dimensional movement and of auxiliary information is proposed for controlling a robotic forceps with voice instructions, where information on the tip-position in three-dimension and on the distance of the previous movement is viewed as the auxiliary one. A degree-related adverb included in the voice instruction is quantified by using fuzzy reasoning, in which such auxiliary information is also applied to update the membership functions so as to realize a user-friendly interface.

1 Introduction

In recent years, robots come to be regarded as very natural in a home, a company, etc., and are utilized for various users. Among them, there have been many cases using a voice interface as means of the communication between human being and robots. Since the voice is the easiest communication approach of human beings, the operation by exclusive input devices, such as a keyboard, is unnecessary, and therefore it can be easily used also for elderly people or the handicapped.

However, in the case that the voice is used for an interface, a certain problem arises. For example, when ambiguous expressions, such as "go to that," are included, the interpretation of "that" depends on each user's subjective scale. Therefore, if a particular setting is carried out for a specific user, then other users may feel inconvenient when using it, without resetting of it. About the interpretation of a voice command, there are a method that interprets a voice command in consideration of a user's viewpoint [1], a method that quantifies the degree-related adverbs contained in a voice command by fuzzy reasoning [2],[3],[4], etc.

In this paper, for the movement control of a robotic forceps operated by the voice that may be used in a laparoscopy, an approach using the image and the auxiliary information on the movement indication in a three-dimensional direction is proposed, together with considering the ambiguity included in the voice instruction. A fuzzy coach-player system [5] is applied as a voice instruction system, using the positional information at the arm tip in three-dimension and the latest movement distance as auxiliary information, where it is aiming at realizing a user-friendly interface by updating membership functions using auxiliary information.

2 Fuzzy Coach-Player System

2.1 Overview of the System

In order for a man and a robot to perform a cooperative task smoothly, mutual intellectual information needs to be exchanged. The fuzzy coach player system used in this research considers the concern between a man and a robot to be a concern between a coach and a player, and takes in the ambiguity included in the voice instruction, the subjective evaluation of the coach, etc.

2.2 Voice Commands from the Coach

Assume that at time step t, an input sequence of fuzzy voice instruction uttered by the coach and collected from a microphone, $v(t)$, is handed over by a voice recognizer, and it can be split into an unnecessary language variable to the motion command, $v_a(t)$, a verb $v_b(t)$, a noun $v_c(t)$, and an adverb (or adverb phrase) $v_d(t)$. Here, $v(k) \in \mathcal{V}$, $v_a(t) \in \mathcal{V}$, $v_b(t) \in \mathcal{V}$, $v_c(t) \in \mathcal{V}$, and $v_d(t) \in \mathcal{V}$, where $\mathcal{V}$ denotes a voice space that represents a time series in signal or character level. Effective language variables to the motion command, $v_b(t)$, $v_c(t)$, and $v_d(t)$ are available to generate a command to the robot.

2.3 System Structure

The structure of an experimental setup in this research consists of a microphone for gathering voice input, a camera for image input, a PC for control, a forceps for grasping an object, and a robot manipulator with 7-dof (called PA10, which is provided by the Mitsubishi Heavy Industries Ltd.) where the forceps is attached on the manipulator tip. The outline of the system is depicted in Fig. 1.

The task is to pick up the object allocated on a cylinder in the work table and convey it to a box. The work coordinate is the same as the base coordinate and the forceps coordinate in the arm tip is shown in Fig. 2.

3 Generation of Voice Commands

3.1 Voice Processing

In the voice processing part, the voice input $v(t)$ coming from the microphone is processed in turn for the voice

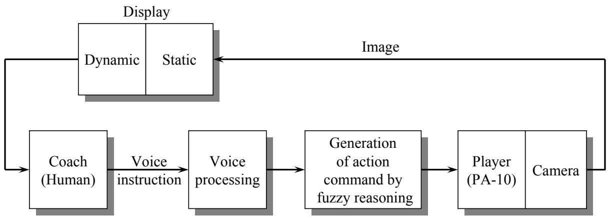


Figure 1: Construction of speech based interface with visual feedback

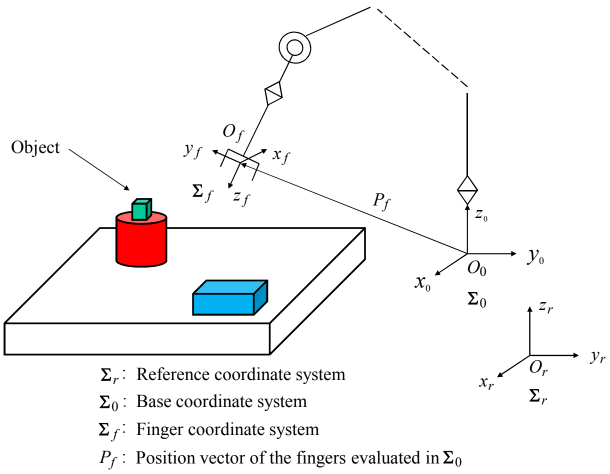


Figure 2: Coordinate system

recognition, the morphological analysis, and the pattern matching in order to generate words $v_b(t)$, $v_c(t)$, and $v_d(t)$ that are related to the motion, the direction, and the degree, respectively.

For the voice recognition, the voice uttered by the user is inputted to the PC through the microphone and then it is transformed into a text by a voice-recognition software, IBM Via Voice. The morphological analysis decomposes the text obtained in the voice recognition into several fundamental words by morphological software, where Chasen, which is a free morphological analyzer for Japanese, was used here. The pattern matching extracts some necessary words by comparing the fundamental words in the text level obtained from the morphological analysis with the keywords tabulated in Table 1.

3.2 Action Generation

The action of the robot arm is generally determined by using $v_b(t)$ and/or $v_c(t)$ obtained in the voice processing. For the actions associated with $v_b(t)$, "hold" and "release" are related to the closing and opening of the forceps, whereas "go" is related to the carrying of an object to the desired position, gripping it. Note that the amount to move is quantified for the forceps by fuzzy reasoning described below, according to $v_d(t)$ included in $v(t)$.

Table 1: Keywords in voice instruction

	Examples
Direction	right, left, up, down, ahead, back
Degree	more, little, very little
Action	go, hold, release

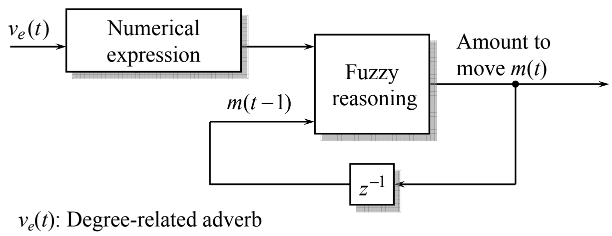


Figure 3: Flow of quantification for degree-related adverbs

The keywords for $v(t)$ are set in Table 1. When "go" is input to $v_b(t)$, the movement direction of the hand tip is determined according to $v_c(t)$. Here, "right" and "left" in $v_c(t)$ are assigned to the direction of x_f-axis in Σ_f, "up" and "down" correspond to the direction of y_f-axis, and "back" and "ahead" are related to the direction of z_f-axis in Σ_f.

4 Fuzzy Rule Extraction Using Auxiliary Information

4.1 Auxiliary Information

The three-dimensional position information on the forceps tip and the amount of movement due to the voice instruction are shown in a display as the auxiliary information giving for the user. It is assumed that such information is composed of a position vector P_f evaluated in the base coordinate.

4.2 Extraction of Fuzzy Rules

4.2.1 Fuzzy Reasoning

The degree-related adverb $v_d(t)$ included in the voice instruction is quantified by fuzzy reasoning, as shown in Fig. 3, where the fuzzy reasoning is assumed to be a simplified one, which consists of two-inputs, one-output, and 16 rules in total.

The first input to the fuzzy reasoning is the voice command $v_d(t)$, where it is labeled and numerically expressed in v_e (i.e., as singleton values) as shown in Table 2. The previous direction of the forceps in x_f, y_f and z_f, i.e., the amount of movement $m_i(t-1)$ [mm] is the second input to the reasoning, and the resultant output is just $m_i(t)$ [mm], where $i = x_f, y_f, z_f$ and the reasoning in each axis is assumed to be performed independently.

Table 2: Numerical expression for degree-related adverbs

Voice command	Very little	Little	Nothing	More
Label	SS	S	M	L
$v_e(t)$	0.2	0.4	0.6	0.8

Table 3: Amount of previous movement

Amount of previous movement	Nothing	Very small	Small	Medium	Large
Label	–	SS	S	M	L
$m_i(t-1)$	0	50	80	150	200

The membership functions for $v_e(t)$, $m_i(t-1)$, and $m_i(t)$ are set as shown in Fig. 4. For the previous and current amounts to move, the corresponding labels and the central (or singleton) values for their membership functions are shown in Table 3 and Table 4 respectively, where the membership functions are assumed to be triangular or singleton, whose values are assumed to be used as the initial ones for the update.

The resultant fuzzy rules is shown in Table 5.

4.2.2 Extraction Method of Fuzzy Rules

It is assumed that in the extraction of fuzzy rules, $a_1, \ldots, a_4$ and $b_1, \ldots, b_4$ in Fig. 4 are updated according to the current position for the forceps tip, the content of the voice instruction, and the amount of the previous movement, keeping the number of fuzzy rules. Here, a_i denotes the center value of each membership function on the support set of $m_i(t-1)$ and b_i denotes the position of the singleton for each membership function on the support set of $m_i(t)$.

Set the initial values for the update of a_i and b_i as given in Table 3 and 4, and update the allocation of all membership functions per each axis by the following method. Note here that in what follows, "the position of the forceps tip is positive" means the case when the value of any axis coordinate takes a positive value, whereas the negative means the case when it takes a negative values. Additionally, "the direction of the voice instruction is positive" implies the case when $v_c(t)$ takes any word out of "right," "up," and "ahead," whereas the negative means the case when it takes any word out of "left," "down," and "back."

1. The update by using the position of the forceps tip and the content of the voice instruction
   - If the position of the forceps tip is positive and the direction of the voice instruction is negative, or the position of the forceps tip is negative and the direction of the voice instruction is positive, then
     $a_i := a_i \times 1.2, \quad b_i := b_i \times 1.2$
   - If the position of the forceps tip is positive and the direction of the voice instruction is positive, or the position of the forceps tip is negative and the direction of the voice instruction is negative, then
     $a_i := a_i \times 0.8, \quad b_i := b_i \times 0.8$
2. The update by using the amount of the previous movement and the degree-related word included in the voice instruction
   - If $m_i(t-1) < 20.0$ and the voice instruction contains "very little," then
     $a_i := a_i \times 0.8, \quad b_i := b_i \times 0.8$
   - If $m_i(t-1) < 20.0$ and the voice instruction contains "more," then
     $a_i := a_i \times 2.0, \quad b_i := b_i \times 2.0$
   - If $30.0 < m_i(t-1) < 50.0$ and the voice instruction contains "very little," then
     $a_i := a_i \times 0.7, \quad b_i := b_i \times 0.6$
   - If $30.0 < m_i(t-1) < 50.0$ and the voice instruction contains "more," then
     $a_i := a_i \times 1.8, \quad b_i := b_i \times 1.8$
   - If $100.0 < m_i(t-1)$ and the voice instruction contains "more," then
     $a_i := a_i \times 1.2, \quad b_i := b_i \times 1.2$

Table 4: Amount of current movement to be output

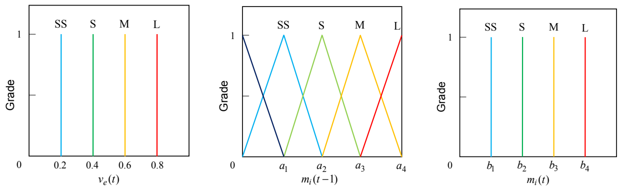
Amount of current movement to be output	Very small	Small	Medium	Large
Label	SS	S	M	L
$m_i(t)$	30	60	120	180

Table 5: Fuzzy rules

		Voice command (degree-related adverbs)			
		SS	S	M	L
Amount of previous movement	SS	SS	SS	SS	S
	S	SS	SS	S	M
	M	SS	S	M	L
	L	S	M	L	L

5 Actual Experiment

In this research, the holding transference of an object controlled by the voice instruction is conducted as actual experiment. The user indicates with the voice, seeing the image provided on the display and taking account of auxiliary information. When it needs to tune the location finely, a forceps is moved to the desired location by following a command, such as "go very little right." The object on the red cylinder is grasped by the indication of "hold" and is



(a) (b) (c)

Figure 4: Membership functions: (a) for the degree-related adverb $v_e(t)$, (b) for the amount of previous movement $m_i(t-1)$, and (c) for the output $m_i(t)$

Table 6: A series of voice instructions

Step number	Command	x_f [mm]	y_f [mm]	z_f [mm]
1	Go more right	60.0	0	0
2	Go right	88.0	0	0
3	Go up	0	108.34	0
4	Go little ahead	0	0	76.80
5	Go very little ahead	0	0	30.17
6	Hold it	0	0	0
7	Go back	0	0	−37.30
8	Go more left	−189.06	0	0
9	Go down	0	−189.06	0
10	Go very little left	0	−72.49	0
11	Release it	0	0	0

conveyed to the blue box, where it is set there after receiving the indication of "release."

As an operational example, a series of performing directional indications is shown in Table 6 where the corresponding updates of the allocation for membership functions are given in Table 7 and Table 8.

6 Conclusions

In this research, it has been demonstrated that in the fuzzy coach-player system using voice instruction, the operation in three-dimensional space became easy by providing the auxiliary information such as the forceps tip position etc., together with using the image information, to the user. For the interpretation of the voice instruction, the user was able to give an instruction such as "go very little right" to the robot, because an ambiguous expression included in the voice instruction was quantified by using a fuzzy reasoning. In addition, the voice instruction became much easier than the conventional method, because the allocation parameters for membership functions were updated by applying auxiliary information so that the fuzzy reasoning was performed using the relevant rules according to the situation.

Table 7: Update of membership allocation for the amount of previous movement

Step number	a_1 [mm]	a_2 [mm]	a_3 [mm]	a_4 [mm]
1	50.00	80.00	150.00	200.00
2	40.00	64.00	120.00	160.00
3	32.00	51.20	96.00	128.00
4	32.00	51.20	96.00	128.00
5	32.00	51.20	96.00	128.00
6	32.00	51.20	96.00	128.00
7	69.12	110.59	207.36	276.48
8	82.94	132.71	248.83	331.78
9	82.94	132.71	248.83	331.78
10	99.53	159.25	298.60	398.13
11	99.53	159.25	298.60	398.13

Table 8: Update of membership allocation for the amount of current movement

Step number	b_1 [mm]	b_2 [mm]	b_3 [mm]	b_4 [mm]
1	30.00	60.00	120.00	180.00
2	24.00	48.00	96.00	144.00
3	19.20	38.40	76.80	115.20
4	19.20	38.40	76.80	115.20
5	19.20	38.40	76.80	115.20
6	19.20	38.40	76.80	115.20
7	41.47	82.94	165.89	248.83
8	49.72	99.53	199.07	298.60
9	49.72	99.53	199.07	298.60
10	59.72	119.44	238.88	300.00
11	59.72	119.44	238.88	300.00

References

[1] Y. Shinyama, T. Tokunaga, and H. Tanaka, "Kairai—Software robots understanding natural language," *J. of Information Processing Society of Japan*, vol. 42, no. 6, pp. 1359–1367, 2001.

[2] H. Kitagami, K. Nishihara, T. Yoshidome, and N. Kawarazaki "A robot operation system based on voice instruction by a quantitative expression," in *Proc. of 2001 JSME Conf. on Robotics and Mechatronics*, 2P2-N8, 2001.

[3] S. Kajiwara, S. Hiratsuka, T. Ishihara, and H. Inooka, "Position control using voice instruction with ambiguous degree expression," *Trans. of the Japan Society of Mechanical Engineers*, C, vol. 69, no. 686, pp. 2735–2742, 2003.

[4] K. Kankubo and S. Kobayashi, "Basic research on application of adverb instruction to robot," Memoirs of the Yuge National College of Maritime Technology, vol. 29, no. 9, pp. 49–55, 2007.

[5] K. Izumi, S. Ishii, and K. Watanabe, "Voice control of a robotic forceps using hierarchical instructions," in *Proc. of 2008 Int. Conf. on Control, Automation and Systems (ICCAS 2008)*, Seoul, Korea, 2008, pp. 810–815.