Proceedings of the 2003 IEEE International Workshop on
Robot and Human Interactive Communication
Millbrae, California, USA, Oct. 31 - Nov. 2, 2003

Robot Position Control via Voice Instruction Including Ambiguous Expressions of Degree

Shinya KAJIKAWA*, Shigeyoshi HIRATSUKA**, Tadashi ISHIHARA*** and Hikaru INOOKA***

*Akita Prefectural University, Honjo, Akita, 015-0055, Japan
**CARC, National Institute of Advanced Industrial Science and Technology, Tokyo Waterfront, Ko-to-ku, Tokyo, 135-0064, Japan.
*** Graduate School of Information Sciences, Tohoku University, Sendai, 980-8579, Japan.
E-mail : kajikawa@akita-pu.ac.jp

Abstract

Ambiguous expressions of degree are frequently used, when instructing another person regarding a task involving movement (,e.g "move it a little" , "lift it more" and so on). To make a more friendly robot, these ambiguous expressions should be quantified adequately by means of a robot control system that offers effective support. In this paper, we aim at constructing an effective controller coping with such ambiguous instructions and making a robot provide useful support for humans. We discuss how to generate appropriate robot arm movement in terms of our sense of distance based on particular expressions of degree. First, we analyzed human arm movement guided by voice instruction including some kinds of expressions of degree. From this analysis, we then obtained a model which estimates a desired length of movement according to expressions of degree in the same manner as does a person. Finally, we executed experiments to show that our model can lead a robot to a desired goal position guided by several instructions.

1 Introduction

Man-machine interface systems using voice information will help us use unfamiliar machine systems, because we can tell the system our requirements easily without employing any operational tools, such as a joystick, keyboard, and so on. Recently, many studies on robot control and robot-human communication via verbal information have been reported[1][2][3][4][5]. These studies focus mainly on generating a control code for robot motion based on voice information or on realizing interactive communication such as conversation. Therefore, the classification of key words extracted from human speech and the analysis of the relationship between them based on grammar are main issues. Usually, the target words are a noun, verb, and predicate.

In natural human conversation, however, many kinds of ambiguous adverbs or expressions of degree are also used to communicate requirements in detail (,e.g. "move it to the left a little", "lift it more" and so on). So that a robot can become a more friendly and smart assistant for us, it should recognize our ambiguous expressions of degree and offer effective support satisfying our requirement. Therefore, the ability to adequately quantify expressions of degree is important in design of an effective voice-controlled robot[6][7].

Fuzzy set theory is well-known as a method that quantifies this kind of ambiguous language[8]. However designing a fuzzy based controller for positioning is difficult task, because repeated voice guidance is needed until the goal is reached and our sense of distance for each expression will be changeable according to the situation, such as the distance remaining to the goal, what instruction was used before, and so on. Thus an efficient membership function is complex and is difficult to design .

In this paper, we aim at constructing a simple but efficient method for controlling the position of a robot arm using degree expressions. In order to design an efficient controller, we must construct a quantification model of expressions of degree whose output matches our sense of distance. For this reason, we analyze human arm movement guided by voice instruction expressing degree and examine the relationship between degree expression and length of movement. From this result, we construct a simple model which estimates suitable length of movement matching a current instruction based on the history of past instructions and movement. Experimental results show that our proposed model can lead a robot to a desired position based on several instructions in the same manner as humans.

2 Voice guided human arm movement

In this section, we investigate human arm movement guided by voice instruction including ambiguous degree expression and analyze the relationship between the length of movement and the expression of degree.

2.1 Experimental setup and method

Figure 1 shows experimental setup. The size of the desk in this figure is 600mm × 900mm (depth × wide). A guide rail, on which a goal position is located, is set in front of an operator.

The operator knows that the goal position is on the guide rail, but does not know the goal's precise position. He moves an object to the right or left along the guide rail under the instructor's voice guidance. The object is a cylindrical shape whose diameter, height, and weight are 60mm, 100mm, and 196g, respectively. The instructor knows the distance between the goal position and the current object's position based on an image from a video camera which is placed in front of the operator. (As shown in Figure 1, goal positions are visible to the instructor only, because those positions are marked on the rail on the opposite side to the operator.) Using this visual information, the instructor informs the operator which direction and how far he should move the object in order to reach the goal position using some kinds of instructions.

In this experiment, eight kinds of instruction in Japanese are used. These instructions include three degree expressions to express the distance to the goal. They are "Sukoshi", "Honno Sukoshi" and "Mou Sukoshi" in Japanese, meaning "a little","just a little", and "a little more" in English, respectively. The instructions are shown in Table.2. In Table.2, we use some symbols which are "**a-l**","**j-a-l**", "**a-l-m**" and "**non**", to easily denote the degree expression used. Symbols "**a-l**","**j-a-l**", and "**a-l-m**" mean "**a** little","just **a** little" and "**a** little more", respectively. And "**non**" means that no degree expression is used.

These instructions are repeated until the position error between the goal and the object is smaller than 3.0mm. During the positioning operation, the distance moved for each instruction is measured. Seven goal positions are set on the right side of the operator. These positions are shown in Table.1. The operators are 14 right-handed males in their twenties.

2.2 Relationship between movement distance and expression of degree

All operators could reach the goal positions after 4 ∼ 7 instructions.This means that the operators seemed to modify the quantification of the degree expressions adequately at each step based on the histories of past movements and the current instruction. Therefore we analyzed the movement distance ratio between the current step and previous one. Figure 2 shows an example of history of the object movement. In this figure, d_{k-1} and d_k indicate the length of movement according to the $k-1_{th}$ and k_{th} instructions, respectively. Here we define a parameter α as the ratio between d_{k-1} and d_k,

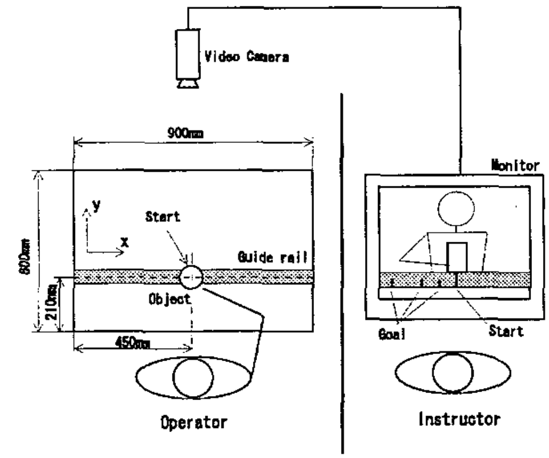

Figure 1: Experimental setup.

Table 1: Distance between start position and goal.

Goal Position (mm)
25, 50, 75, 100, 125, 150, 175

$$\alpha_{+(-)} = \left|\frac{d_k}{d_{k-1}}\right|, \quad (1)$$

where

α_+ : the direction of hte $k-1_{th}$ and k_{th} movements are same (forward movement)

α_- : the direction of the $k-1_{th}$ and k_{th} movements are not same (reverse movement).

The average values of parameters α_+ and α_- are sumarized as shown in Table 3 and Table 4, respectively. In these tables, the column indicates the expression of degree used at the $k-1_{th}$ step and the row is that of the k_{th} step. Symbol "-" denotes that the combination of degree expressions corresponding to this cell is not used in the experiment.

From this result, we notice that α_+ and α_- have various values according to the combination of instructions at $k-1_{th}$ and k_{th}. Furthermore, it is important that α_- is always smaller than 1.0 for all combinations. This means that the object's position

Table 2: Voice instruction with degree expression.

Symbol	Expression in Japanese / Expression in English
non	Hidari (Migi) ni ugokashite
	Move it to the left (right)
a-l	Hidari (Migi) ni "sukoshi" ugokashite
	Move it to the left(right) "**a** little"
j-a-l	Hidari (Migi) ni "honno sukoshi" ugokashite
	Move it to the left(right) "just **a** little"
a-l-m	Hidari (Migi) ni "mou sukoshi" ugokashite
	Move it to the left(right) "**a** little more"

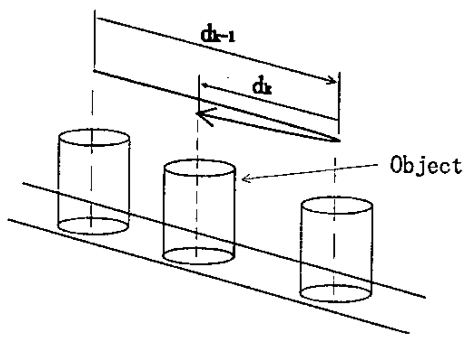

Figure 2: Definition of the length of movement d_k.

converges to a certain point after several repetitions of reverse movements. Therefore we can infer that the instructor can lead the object to its goal position by repeating instruction for reverse movement after the operator passed by the goal.

3 Degree expression - Movement distance Model

Here we propose simple model which quantifies the degree expression using the results shown in previous section. We define two movment distance ratio matrices, α_+ and α_-, based on the values shown in Table.3 and Table.4 as follows:

$$\alpha_+ = \begin{pmatrix} 0.96 & 0.18 & 0.10 & 0.16 \\ 1.00 & 1.02 & 0.31 & 0.98 \\ 1.00 & 0.97 & 1.35 & 1.31 \\ 1.00 & 1.13 & 0.49 & 1.04 \end{pmatrix} \quad (2)$$

$$\alpha_- = \begin{pmatrix} 0.50 & 0.25 & 0.50 & 0.50 \\ 0.50 & 0.40 & 0.45 & 0.49 \\ 0.50 & 0.50 & 0.71 & 0.90 \\ 0.50 & 0.50 & 0.68 & 0.61 \end{pmatrix}, \quad (3)$$

where 1.00 and 0.50 in α_+ and α_- are adequate selected values, because these cells in Table 3 and Table 4 are empty. According to the combination of the k_{th} and $k-1_{th}$ instructions, the robot chooses the movement distance ratio from the elements of

Table 3: Ratio of movement distance according to combination of degree expression : α_+ (movement direction : right → right or left → left).

α_+		k_{th} instruction			
		non	**a-l**	**j-a-l**	**a-l-m**
$k-1_{th}$ inst-ruction	**non**	0.96	0.18	0.10	0.16
	a-l	-	1.02	0.31	0.98
	j-a-l	-	0.97	1.35	1.31
	a-l-m	-	1.13	0.49	1.04

Table 4: Ratio of movement distance according to combination of degree expression : α_- (movement direction : right → left or left → right).

α_-		k_{th} instruction			
		non	**a-l**	**j-a-l**	**a-l-m**
$k-1_{th}$ inst-ruction	**non**	-	0.25	-	-
	a-l	-	0.40	0.45	0.49
	j-a-l	-	-	0.71	0.90
	a-l-m	-	-	0.68	0.61

these matrices, and estimates the movement distance d_k by Eq.(4) and Eq.(5),

$$d_k = \phi_k d_{k-1} (k \geq 2) \quad (4)$$

$$\phi_k = \begin{cases} \alpha_+(i,j) & (forward\ movement) \\ \alpha_-(i,j) & (reverse\ movement) \end{cases} \quad (5)$$
$$(1 \leq i, j \leq 4),$$

where i and j indicate the column and row of α_+ and α_-, respectively, and they point out the combination of the degree expression used in the k_{th} and $k-1_{th}$ instructions. The robot position at the k_{th} instruction is given by following equations.

$$X_k = \begin{cases} d_1 & (k=1) \\ X_{k-1} + d_k & (k \geq 2) \end{cases} \quad (6)$$

X_1 or d_1 should be given adequately beforehand.

3.1 Modification of movement distance

Figure 3 shows the situation where the robot reaches X_k passing by the goal position X_{goal} after the k_{th} instruction with the instructor then requiring reverse movement at the $k+1_{th}$ step. At this time, we can easily infer that the goal position exists between X_{k-1} and X_k. With this knowledge, we can restrict the goal area as follows:

Case 1: Movement direction: right → left

$$X_{k-1} < X_{goal} < X_k$$

Case 2: Movement direction: left → right

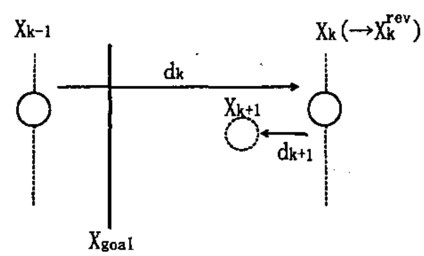

Figure 3: Existance area of goal position.

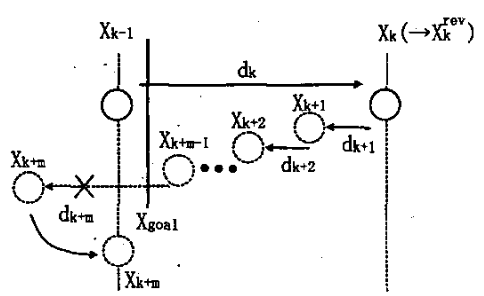

Figure 4: Modification of movement distance.

$$X_k < X_{goal} < X_{k-1}.$$

Figure 4 shows the situation where, after m times movement from X_k, the robot jumps over position X_{k-1}; the border of the area in which the goal exists. In this case, we modify the movement distance d_{k+m} so as to stop the robot movement at X_{k-1}. This modification law is described as follows:

$$d_{k+m} = X_{k-1} - X_{k+m-1} \quad (m > 1). \tag{7}$$

By introducing this modification, we can decrease the number of instructions necessary to reach the goal position.

4 Experiment

The experimental system is shown in Figure 5. This system consists mainly of a robot manipulator (RA-1, Mitsubishi Electric Corp.) and the dictation engine of the speech recognition system "Via-Voice (IBM)" for the Japanese language.

The instructor guides this robot to the goal position using voice instruction. The instructor watches the current robot position and selects a suitable instruction from the eight expressions shown in Table 1. Via-Voice extracts and recognizes two key

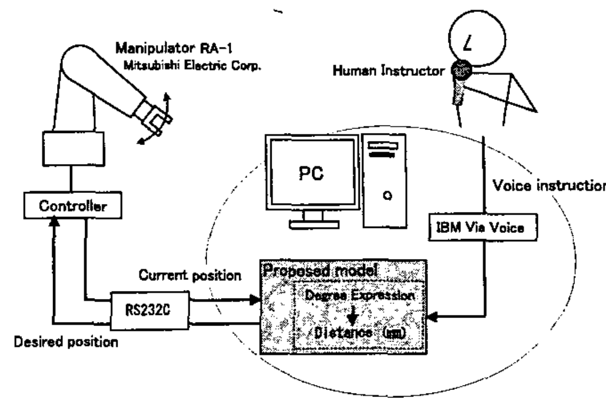

Figure 5: Experimental System.

Table 5: Model decides movment distance ratio ϕ

	Movement distance ratio ϕ
Model A	0.91 or 0.60
Model B	α_+ or α_-

words from the inputted instruction; the one is a degree expression concerning movement distance ("a little","just a little", "a little more") and the other stands for the intended direction of movement ("right" or "left"). Based on this recognition, the distance ratio of movement ϕ is chosen from the elements of α_+ or α_- defined by Eq.(2) and Eq.(3), respectively, and distance the robot should move is determined. The desired length of movement and the command regarding direction are sent to the robot controller via RS232C.

In this experiment, we prepared another, simpler model to determine movement distance ratio ϕ. This model is expressed as follows:

$$\phi_k = \begin{cases} 0.91 & (forward\ movement) \\ 0.60 & (reverse\ movement). \end{cases} \tag{8}$$

The values, 0.91 and 0.60 are the averages of all of human arm movement ratios for forward and reverse movements, respectively. We call this model Model A, and the one using $\alpha_{+(-)}$ Model B. In this method, the distance of first movement d_1 must be prepared beforehand according to the kind of instruction. Using the observed data of the distance of human arm movement measured in Section 2, we determined the initial value d_1 as shown in Table 6. These values are averages for each expression used in the first instruction.

We conducted two experiments here. One, **Exp.A**, was executed to compare the motion generated by Model A with that of Model B. The other, **Exp.B**,

Table 6: Movement distance for first instruction

	non	a-l	j-a-l
Distance (mm)	249.5	155.6	75.0

Table 7: Goal position

	distance (mm)	direction
Exp.A	100, 175	right hand
Exp.B	200	right hand

demonstrates the effectiveness of the method for modifying the length of movement, which was explained in Section 3.1.

The goal positions set for each experiment are shown in Table 7. In **Exp.A**,two kinds of goal positions were established to the right of the robot. There are 100mm or 175mm from the robot's initial position. As for **Exp.B**, the goal is set 200mm to the right of the robot.

5 Experimental Results

5.1 Exp.A

Figure 6 and Figure 7 show that the process of movement toward the goal position 100mm and 175mm, respectively. In these figures, a typical example of a human operator's movement is also plotted for comparison with our method. Markers □, ●, and ○ represent the results of Model A, Model B, and a human operator, respectively. Table 8 and Table 9 indicate the sequence of instructions used in each experiment. In these tables, "L" and "R" inside () denote the direction of movement, left or right, respectively.

Comparing these results, Model A can reach the goal position within 11 instructions as did the human operator. Model B also can achieve accurate positioning. This means that if the movement distance ratio for reverse motion ϕ is smaller than 1, the robot can reach an unknown goal position by only voice instruction. The number of instructions could be more efficiently reduced by defining suitable value of ϕ for each expression as Model A.

However, the details of the process of converging to the goal position differ between the robot motion of Model A (or Model B in Figure 6) and the human one. Human approaches the goal more gradually than does the robot and restricts changes in the direction of movement. Human motion consists mainly of forward movement. These characteristics give us the impression that human motion is more sensitive than that of a robot. In order to make the robot motion more closely resemble that of humans, we have to define a different movement distance ratio ϕ for each instruction step. At least, the ratio at the 2nd step should be smaller, because the length of robot's movement at this step is longer than that of human and is used for deciding the length of later movement.

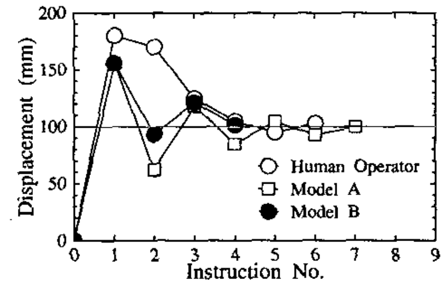

Figure 6: Experimental results (x_{goal} = 100mm).

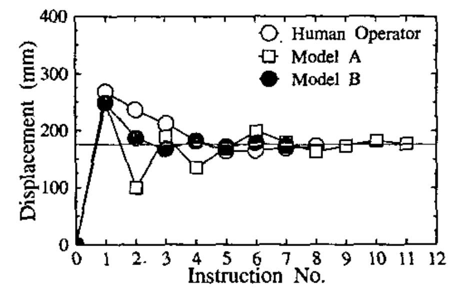

Figure 7: Experimental results (x_{goal} = 175mm).

5.2 Exp.B

Figure 8 illustrates the experimental results. In this figure, Model C shows the case in which Model B is used to decide the next position but no modification of the length of movement is applied, even if the robot passes over the goal area. Marker * denotes the time point at the modification should be performed. In this experiment, the modification of movement distance should be applied at the $6th$ and $9th$ instructions. Model B executes the modification of movement at the $6th$ instruction, so after that it can reach the goal position immediately. On the other hand, Model C needs 12 instructions. From this comparison, we confirm the effectiveness of the modification of the length of movement.

6 Conclusion and Future works

In this paper, we propose a simple and effective voice-control method applied to a positioning task performed by a robot. Especially, we aim at designing a controller which can understand an ambiguous expression of degree in the instructions used for describing the distance to a goal, such as "move it to the left a little or a little more " and so on. In or-

Table 8: *Order of instruction with degree expression (x_{goal}=100mm).*

	Human	Model A	Model B
1st	a-l (R)	a-l (R)	a-l (R)
2nd	a-l (L)	j-a-l (L)	a-l (L)
3rd	a-l-m (L)	j-a-l (R)	j-a-l (R)
4th	a-l-m (L)	j-a-l (L)	j-a-l (L)
5th	j-a-l (L)	j-a-l (R)	–
6th	j-a-l (R)	j-a-l (L)	–
7th	–	j-a-l (R)	–

Table 9: *Order of instruction with degree expression (x_{goal}=175mm).*

	Human	Model A	Model B
1st	non (R)	non (R)	non (R)
2nd	a-l (L)	a-l (L)	a-l (L)
3rd	a-l-m (L)	a-l (R)	j-a-l (L)
4th	a-l-m (L)	j-a-l (L)	j-a-l (R)
5th	j-a-l (L)	j-a-l (R)	j-a-l (L)
6th	j-a-l (R)	a-l-m (R)	j-a-l (R)
7th	j-a-l (R)	j-a-l (L)	j-a-l (L)
8th	j-a-l (R)	j-a-l (L)	–
9th	–	j-a-l (R)	–
10th	–	j-a-l (R)	–
11th	–	j-a-l (L)	–

der to realize the adequate quantification for these expressions, we analyze human arm movement as guided by them and extract the relationship between the history of sequential instructions and that of the length of movement. Based on this relationship, we define a simple model to estimate desired length of movement for the combination of a current expression and those used one step previously. This model is very simple but works efficiently to guide the robot to a desired goal.

We will develop this method to operate a robot in 2D or 3D space by adding various combination of the direction command and degree expressions.

References

[1] J.Ido, K.Takemura, Y.Matsumoto and T.Ogasawara, "Robotic Receptionist ASKA : A Reserch Platform for Humam-Robot Interaction", Proc. of IEEE International Workshop on Robot and Human Interactive Communication, 306-311.2002.

[2] O.Rogalla, M.Ehrenmann, R.Zollner, R.Becher and R.Dillmann, "Using Gesture and Speech Control for Commanding a Robot Assistant", Proc. of IEEE International Workshop on Robot and Human Interactive Communication, 454-459,2002.

[3] S.Takata, S.Kawato and K.Mase, "Conversational Agent Who Achieves Tasks While Interacting with Humans Based on Scenarios", Proc. of IEEE International Workshop on Robot and Human Interactive Communication, 235-239,2002.

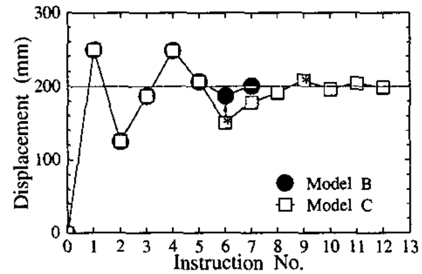

Figure 8: *Comparison of two experimental results by Model B and Model C (x_{goal}=200mm).*

Table 10: *Order of instruction with degree expression (x_{goal}=175mm).*

	Model B	Model C
1st	non (R)	non (R)
2nd	j-a-l (L)	j-a-l (L)
3rd	a-l (R)	a-l (R)
4th	a-l-m (R)	a-l-m (R)
5th	j-a-l (L)	j-a-l (L)
6th	j-a-l (L)	j-a-l (L)
7th	j-a-l (R)	a-l (R)
8th	–	j-a-l (R)
9th	–	j-a-l (R)
10th	–	j-a-l (L)
11th	–	j-a-l (R)
12th	–	j-a-l (L)

[4] S.Lauria, G.Bugmann, T.Kyriacou, J.Bos and E.Klein, " Converting Natural Language Route Instructions into Robot Executable Procedures", Proc. of IEEE International Workshop on Robot and Human Interactive Communication, 223-228,2002.

[5] K.Kitamura and C.Asakawa, " "Speech Pointer" A Non-visual User Interface Using Speech Recognition and Synthesis", Proc. of IEEE International Conference on Systems, Man and Cybernetics, III,1083-1088,1999.

[6] N.Tejima, "Factors Influencing Drivability for a Voice-Controlled Wheelchair ", Proc. of World Congress on Medical Physics and Biomedical Engineering ,FR-Ba204-03,2000.

[7] S.Hiratsuka, S.Kajikawa and H.Inooka, "Analysis of Ambiguous Adverbial Expression Used for Instruction of Positioning Control", Proc. of International Conference on Control, Automation and Systems,186-189,2001.

[8] L.A.Zadeh, "Fuzzy Set", Information and Control, **8**, 338-353,1965.