Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2005). pp.199-202. 2005.

ACTION CONTROL DIALOGUE CORPUS

Kotaro FUNAKOSHI

Department of Computer Science,
Tokyo Institute of Technology,
Oookayama 2-12-1 Meguro, Tokyo
koh@cl.cs.titech.ac.jp

Kaisei HAMAMOTO

Department of Computer Science,
Tokyo Institute of Technology,
Oookayama 2-12-1 Meguro, Tokyo
hamamoto@cl.cs.titech.ac.jp

Naoko KURIYAMA

Department of Human System Science,
Tokyo Institute of Technology,
Oookayama 2-12-1 Meguro, Tokyo
kuriyama@hum.titech.ac.jp

Takenobu TOKUNAGA

Department of Computer Science,
Tokyo Institute of Technology,
Oookayama 2-12-1 Meguro, Tokyo
take@cl.cs.titech.ac.jp

ABSTRACT

Action control dialogue is a kind of task-oriented dialogue in which a commander controls the action of other agents through verbal interaction. In this paper, we introduce an action control dialogue corpus and also a Wizard-of-Oz tool to collect dialogue data. The corpus contains 35 video-recorded action control dialogues in which a pair of participants: a commander and a follower collaboratively solve a Tangram puzzle. The tool is a derivative of a spoken dialogue system with which users can command software robots in a virtual world. It uses a mouse-based interface instead of the original speech interface that enables smooth manipulation of robots.

1. INTRODUCTION

Non-industrial or domestic robots have come to the fore in recent years. Language is an important communication tool to interact with such robots. In particular the technology enabling action control of robots through verbal interaction will be a key factor for robots to become widely used. We call such interaction *action control dialogue*.

A small Japanese corpus of dialogue between humans and robots is available [1], but it is not enough to explore various linguistic phenomena taking place in action control dialogue. Therefore, we have been working on collection of additional action control dialogue data.

This paper reports two products of our activity. One is a video-recorded corpus of action control dialogues between humans working on Tangram puzzles. The other is a tool for simulating human-machine action control dialogue.

In section 2, we explain the notion of action control dialogue. In section 3 and section 4, we introduce the action control dialogue corpus that we collected and the tool for action control dialogue simulation in the Wizard-of-Oz method respectively.

2. ACTION CONTROL DIALOGUE

In action control dialogue, a commander controls the action of other dialogue participants (followers). In case of dialogue systems such as SHURDLE [2] and $\mathcal{K}2$ [3] (see Fig.1), a user takes the role of the commander and the system takes the role of the follower(s). The system can consist of one or more agents. Here, agents can be either hardware robots or software robots in a virtual world.

Action control dialogue is a kind of *command & control dialogue*. Command & control dialogue also covers the control of non-physical acts by agents such as meeting room reservation, but action control dialogue focuses on the con-

Fig. 1. $\mathcal{K}_2$

trol of physical acts (action) along the continuous time-line.

Although these dialogues are basically *master-slave*, they also allow for *mixed-initiative*. A follower waits for an instruction from the commander rather than suggesting a next action. Furthermore, he can initiate any communicative action. He can request an order and ask the commander a question at any time.

3. TANGRAM TASK CORPUS

In this section, we report an action control dialogue corpus, Tangram Task Corpus, which consists of video recording dialogues between pairs of participants who collaboratively solve a Tangram Puzzle.

3.1. Tangram Puzzle

Tangram is a combination puzzle consisting of seven geometric pieces. Given only the contours of objective figures, the goal of the puzzle is finding out the combinations of puzzle pieces that forms the target figure. Figure 2 shows an example of a question and figure 3 shows its answer. A given goal may allow for several different combination patterns.

3.2. Experiment Settings

Participants

Each dialogue includes a commander given an target figure and a follower given a set of Tangram puzzle. The commander can watch the puzzle in front of the follower, and command the follower verbally to move pieces. The follower combines puzzle pieces according to the instructions

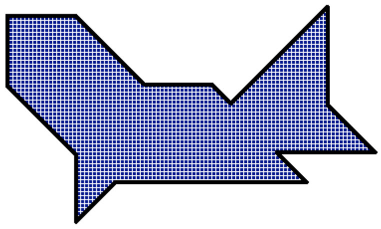

Fig. 2. Tangram Puzzle: Goal

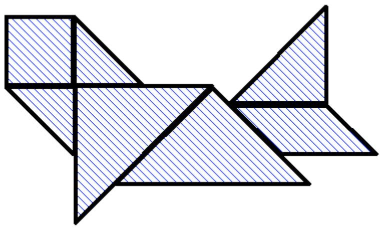

Fig. 3. Tangram Puzzle: Answer

of the commander but cannot see the objective figure that is given only to the commander. Most participants are native Japanese but some are non-Japanese.

Conversation Restrictions

We imposed no restriction on conversations between commanders and followers as far as they concerned the puzzle. Thus, both participants could make any question and suggestion to the other participant. Followers could even move puzzle pieces without instructions from commanders. 30 minutes time limit was set to solve each puzzle. All the dialogues were in Japanese.

Recording

We recorded the dialogues between commanders and followers in video and audio. The workbench (see Fig.4) where a follower manipulates puzzle pieces and the faces of the commander and the follower were recorded with three video cameras in sync. The voices of two participants were recorded as separate channels (left and right) of one stereo audio stream also in sync with video.

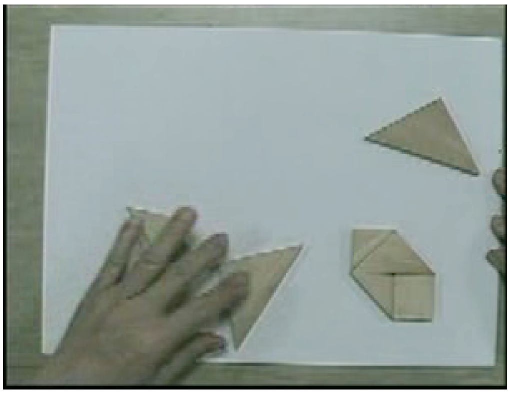

Fig. 4. Snapshot of the Workbench

R: えっとその せ 平行四辺形あるじゃないですか
well that there is a parallelogram right?
L: うん
yeah
R: ええ
yeah
R: それをその今う そこにスライドさしたところの右っ側に
(put) it to the right of there the place (you) slided (a thing) to
L: ふん こっち
well here
R: いや そっちじゃなくて
no not that way
L: え こっち
oh here
R: そっちじゃなくて その
no not that way that
L: こ こ こっち？
he he here?
R: そっち側で
there

Fig. 5. A fragment of the corpus

Situations

We used two different recording situations, that is, face-to-face and non-face-to-face. In both situations, participants hear raw voice of each other in the same room.

face-to-face: The commander and the follower face each other across the workbench. Thus, participants can watch the faces of each other, but they observe the puzzle from opposite directions.

non-face-to-face: The commander and the follower sit with their backs to each other. The commander observes the workbench on video as shown in Fig.4. Thus, participants share almost the same view of the puzzle.

3.3. Results

In total, 35 dialogues including 3 trials between native Japanese spakers and 4 trials between native Japanese speakers and non-native Japanese speakers were recorded. The recording time is around in 8 hours[1]. We used 37 subjects where each pair subject were two familiar with each other. 14 dialogues were in face-to-face and 21 dialogues were in non-face-to-face setting.

All speech has been transcribed in draft form with no formal rule of utterance delimitation. Current delimitation relies on the intuition of transcribers. A fragment of tracscription is shown in Fig.5.

The corpus is available from the authors under certain restrictions.

[1]The intervals of silence are included.

4. $\mathcal{K}3$-WOZ

The Tangram Task Corpus reported in section 3 can be used to discover interesting phenomena in action control dialogue. However, the collected dialogues are very complex, thus it is often difficult to directly apply the corpus to evaluations of simple discourse models in early stages of development. Therefore, we developed a tool to simulate human-machine dialogues that would be less complex than human-human unconstrained dialogues.

4.1. Wizard-of-Oz

Wizard-of-Oz or WOZ, originated as a famous story, is the name of a method to collect dialogue data or evaluate dialogue systems [4]. This method can be used to collect dialogue data in the domain of a target dialogue system to clarify what kinds of linguistic expressions and phenomena must be handled by the system before it is built.

In the WOZ method, a human hiding in back of a dummy of a target dialogue system simulates the behavior of the dialogue system. Subjects are generally instructed to accomplish a task (e.g., room reservation) by talking with a dialogue system developed to assist the task.

4.2. Tool Overview

The tool, $\mathcal{K}3$-WOZ, is based on the spoken dialogue system, $\mathcal{K}3$, the successor of $\mathcal{K}2$ system in which users can command software robots in a virtual world [3]. These robots can walk around and move objects in a virtual world. $\mathcal{K}3$-WOZ provides a mouse-based interface rather than the original speech interface.

The overview of the experimental configuration using $\mathcal{K}3$-WOZ is shown in Fig.6. A subject commands robots shown in her video display (see Fig.7) via speech, but in reality these robots are manipulated by the wizard monitoring and interpreting the subject's speech.

The wizard monitors another video display (see Fig.8). It has an extra view of the virtual world through which he can control the robots using mouse operations. Keyboard shortcuts are also available. Besides the objects, any points on the floor are also selectable as objects and destinations of robots' actions with mouse clicks on the view. Therefore with a little training, a wizard would manipulate robots as quickly as speech.

$\mathcal{K}3$-WOZ can be used not only for corpus collection but also for psycho-linguistic experiments and so on.

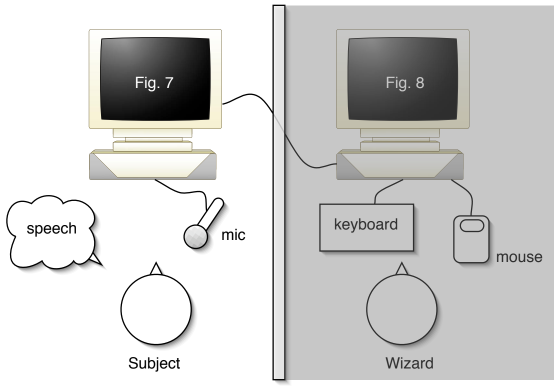

Fig. 6. $\mathcal{K}3$-WOZ: Overview

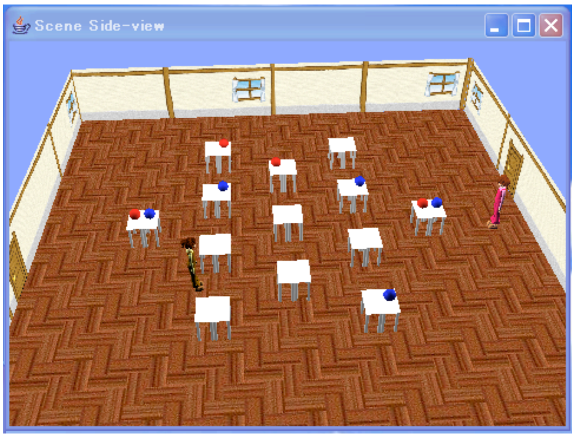

Fig. 7. $\mathcal{K}3$-WOZ: Subject's view

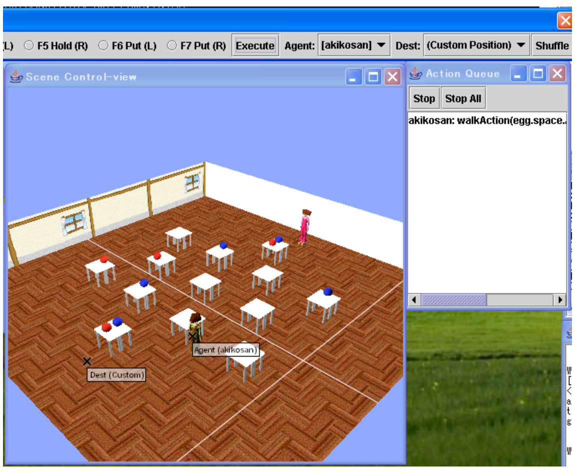

Fig. 8. $\mathcal{K}3$-WOZ: Wizard's view

5. ACKNOWLEDGEMENT

We thank to all subjects who participated in the corpus collection and to KOYAMA Tomofumi who helped in the development of $\mathcal{K}3$-WOZ.

6. REFERENCES

[1] Tanaka-Tokunaga Laboratory, "Quasi Dialogue Corpus," 2001, http://www.cl.cs.titech.ac.jp/qdc/.

[2] Terry Winograd, *Understanding Natural Language*, Academic Press, 1972.

[3] Takenobu Tokunaga, Kotaro Funakoshi, and Hozumi Tanaka, "K2:animated agents that understand speech commands and perform actions," in *The 8th Pacific Rim International Conference on Artificial Intelligence (PRICAI2004)*, 2004.

[4] Dafydd Gibbon, Inge Mertins, and Roger K. Moore, Eds., *Handbook of Multimodal and Spoken Dialogue systems*, Kluwer Aaademic Publishers, 2000.