Conversational Animated Agent System K3

Kotaro Funakoshi Takenobu Tokunaga Hozumi Tanaka
Department of Computer Science, Tokyo Institute of Technology
Ookayama 2-12-1, Meguro, Tokyo 174-0062, Japan
+81-3-5734-2685, {koh,take,tanaka}@cl.cs.titech.ac.jp

1. INTRODUCTION

We have seen significant developments in technology of speech recognition and natural language processing in recent years. Major breakthroughs in the area of computer graphics have enabled us to generate complex, yet realistic 3-D animated agents in a virtual environment. Researchers are now in a good position to go beyond SHRDLU [4] by combining these technologies [2]. We present a conversational animated agent system, $\mathcal{K}3$. A screen shot of $\mathcal{K}3$ is shown in Fig. 1.

Since all the actions carried out by agents of the $\mathcal{K}3$ system are visualized, we can evaluate the system performance by observing its animation. Visualizing the agents' actions sheds light on many interesting issues from a cognitive science point of view; more complex processes should be considered than those discussed in most conventional natural language understanding systems.

Figure 1: Screenshot of $\mathcal{K}3$

2. SYSTEM OVERVIEW

The architecture of $\mathcal{K}3$ is illustrated in Fig. 2. The current system accepts simple Japanese utterances which command agents to perform physical actions, such as "Take the blue ball." Utterances including anaphora and ellipses, and utterances repairing agents' misunderstanding can be accepted.

The speech recognition module receives the user's speech input and generates a sequence of words. The syntactic/semantic analysis module analyzes the word sequence to extract a case frame. This module accepts ill-formed speech input including postposition omission, inversion, and self-correction [1]. At this stage, not all case slots are necessarily filled, because of ellipses in the utterance. Even in cases where there is no ellipsis, instances of objects are not identified at this stage.

Resolving ellipses and anaphora, and identifying instances in the world is performed by the discourse analysis module. Anaphora resolution and instance identification are achieved by using plan-knowledge [3]. When instance identification fails, the agent asks a clarification question to the user and interprets the user's answer.

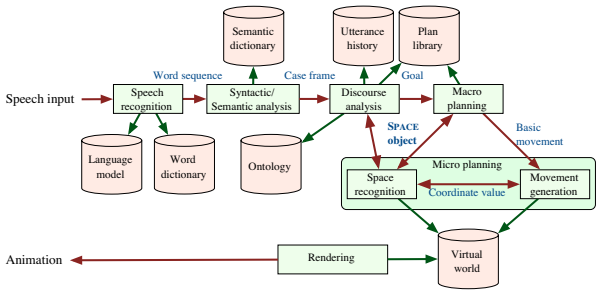

Figure 2: System Architecture

The discourse analysis module extracts the user's goal as well and hands it over to the planning modules, which build a plan to generate the appropriate animation. In other words, the planning modules translate the user's goal into animation data. However, the properties of the user's goal and animation data are very different and straightforward translation is rather difficult. The user's goal is represented in terms of symbols, while the animation data is a sequence of numeric values. To bridge this gap, we take a two-stage planning approach – macro- and micro-planning.

In the macro-planning stage, the planner needs to know the physical properties of objects, such as their size, location and so on. For example, to pick up a ball, the agent first needs to move to the location at which he can reach the ball. In this planning process, the distance between the ball and the agent needs to be calculated. This sort of information is represented in terms of coordinate values of the virtual space and handled by the micro-planner.

To interface the macro- and micro-planning, we introduced a hybrid representation of space [3] which simultaneously represents both the symbolic and numeric nature of a location.

3. REFERENCES

[1] K. Funakoshi, T. Tokunaga, and H. Tanaka. Processing Japanese self-correction in speech dialog systems. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, 2002.

[2] H. Tanaka, T. Tokunaga, and Y. Shinyama. Animated agents capable of understanding natural language and performing actions. In *Life-Like Characters*. Springer, 2004.

[3] T. Tokunaga, K. Funakoshi, and H. Tanaka. K2: Animated agents that understand speech commands and perform actions. In *Proceedings of the 8th Pacific Rim International Conference on Artificial Intelligence*, 2004.

[4] T. Winograd. *Understanding Natural Language*. Academic Press, 1972.