In Proceedings of the European Workshop on Service and Humanoid Robots (ServiceRob '2001), Santorini, Greece, 25-27 June 2001.

HUMAN-ROBOT INTERACTION BASED ON SPOKEN NATURAL LANGUAGE DIALOGUE

Dimitris Spiliotopoulos, Ion Androutsopoulos and Constantine D. Spyropoulos

Software and Knowledge Engineering Laboratory
Institute of Informatics and Telecommunications
National Centre for Scientific Research "Demokritos"
P.O. Box 60228, Ag. Paraskevi 153 10, Athens, Greece
email: {dspiliot, ionandr, costass}@iit.demokritos.gr

ABSTRACT

We report on recent work on human-robot spoken dialogue interaction in the context of Hygeiorobot, a project that aims to build a mobile robotic assistant for hospitals. Spoken dialogue systems are particularly suitable to this context, as the robot does not carry a keyboard or other common interaction devices, and is intended to be used by people with little or no computing experience. In this paper, we concentrate on dialogue management issues. After providing a brief survey of dialogue management techniques, we focus on particular issues that need to be addressed in human-robot interaction, and the considerations that influenced the design of Hygeiorobot's dialogue manager. We then describe in detail the spoken dialogue capabilities of Hygeiorobot's current demonstrator and the tasks that the demonstrator can perform, concluding with plans for future work.

Keywords: spoken dialogue systems, human-computer interaction, robots.

1 Introduction

Robots in current use perform mostly tasks that require little, if any, interaction with casual users; for example, heavy load continuous work in factories, power plants, etc. In recent years, however, robotic assistants are becoming more common in environments such as offices or houses, where robots often need to communicate with users much less exposed to technology than their previous operators. In situations of this type, communication via spoken dialogue systems (SDSs) appears to be a promising approach.

SDSs allow users to interact with machines by means of spoken dialogues in natural language. The general architecture of SDSs comprises six components, as shown in Figure 1. The speech input is first processed by a speech recognizer, which converts it to a written form. This is then passed to the language analyzer, which constructs a logical representation of the user's utterance. Using this representation, information on the previous discourse, and knowledge of the task to be performed, the dialogue manager may then decide to communicate with an external application or device, in our case the robot's controller, or convey a follow-up message to the user. In the latter case, a logical representation of the message is passed to response generator, which generates an appropriate response in written form and passes it to the speech synthesizer.

This paper focuses mostly on dialogue management, based on work performed in the context of "Hygeiorobot", a project whose goal is to develop a mobile robotic assistant for hospitals.[1]

[1] Hygeiorobot is a project funded by the Greek Secretariat of Research and Technology. The project's consortium consists of ICCS-NTUA (coordinator), NCSR "Demokritos", and the University of Piraeus.

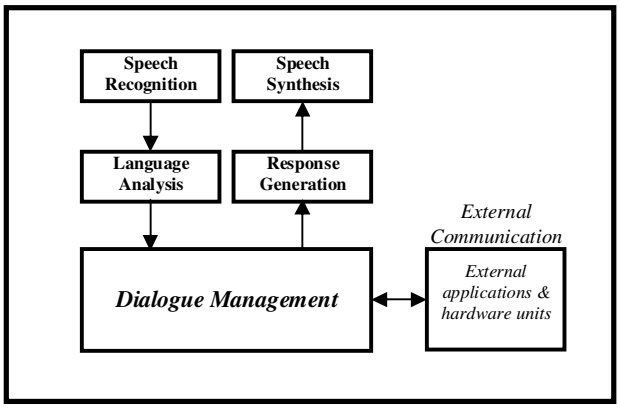

Figure 1: Architecture of spoken dialogue systems

The robot is intended to perform simple tasks in hospitals, such as the delivery of messages or medicines to particular rooms, interacting with hospital staff via spoken dialogues.

Section 2 below provides a brief introduction to spoken dialogue management techniques. Section 3 discusses previous work on natural language interaction with robots. Section 4, then, describes the SDS capabilities of Hygeiorobot. Section 5 concludes and provides directions for future work.

2 Dialogue management techniques

The most commonly used and simplest dialogue management techniques are state-based [AO1995, McT1997, McT1998]. These techniques represent the possible dialogues by a series of states, as shown in Figure 2. At each state, the system may ask the user for specific information, it may generate a response to the user, or it may access an external application. The structure of the dialogue is predefined, and at each state the user is expected to provide particular inputs. This makes the user's utterances easier to predict, leading to faster development and more robust systems at the expense of limited flexibility in the structure of the dialogues.

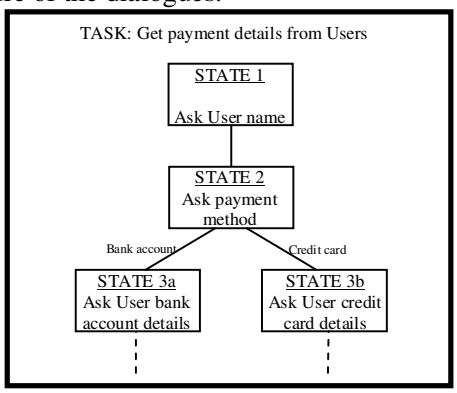

Figure 2: State-based dialogue

For simple tasks, state-based techniques are often the most practical solution. In complex tasks, however, state graphs become extremely large and difficult to maintain, and they lead to long dialogues that users may find irritating.

Frame-based techniques use frames instead of series of states [HStD+1996, VvZ1996], as shown in Figure 3. In this case, each frame represents a task or subtask, and it has slots representing the pieces of information that the system needs in order to complete the task. The system formulates questions to fill in particular slots that remain empty (e.g. the day slot in Figure 3), but the user may get the initiative of the dialogue and provide more information than asked (e.g. both the day and month). This additional information is used to fill in more slots, saving the user from having to answer subsequent questions, and leading to shorter dialogues compared to state-based approaches. On the other hand, user utterances become less restricted and, hence, harder to predict, compared to state-based techniques, which increases the time needed to develop a robust system.

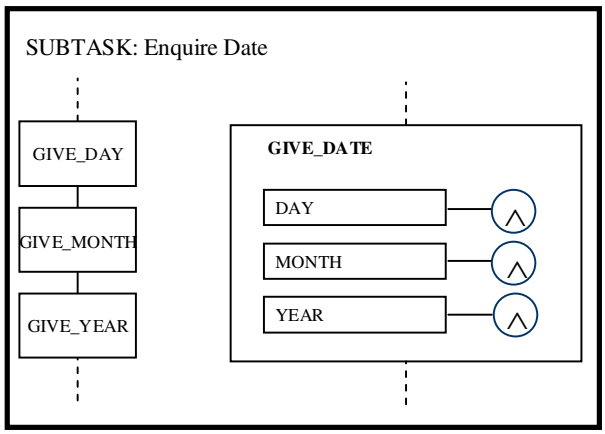

Figure 3: Frame-based dialogue

Rather than modeling the task, plan-based techniques concentrate on identifying the user's plan and determining how they can contribute towards the execution of that plan [FA1993, ASF+1994, AMRS1996, ABD+2001]. This is a dynamic process, whereby new information from the user may force the system to modify its initial perception of the user's plan and its possible contribution. Plan-based techniques typically allow for greater degrees of user initiative in the dialogues, compared to previously mentioned approaches, and have proven to be particularly well suited to problems where the pieces of information or actions that are needed to perform a task are hard to predict in advance (e.g. repairing a machine, rather than simply accessing a bank account). The implementation and maintenance of plan-based systems, however, is far more complex, compared to systems based on the previous approaches.

3 Natural language and robots

There have been several attempts to build mobile robots with natural language interaction capabilities, though the language facilities of many of them were rather simple, and would not qualify as full SDSs. We have studied several of them, in order to identify particular issues that need to be taken into consideration when developing SDSs for mobile robots:

RHINO is a robotic guide that can move within a museum and describe particular exhibits [BCF+1998]. It does not support true dialogues, but can recognize simple phrases like "execute tour number 3". Tours, then, follow fixed routes, using canned spoken utterances. An older robotic guide was Polly, a vision-based robot that could offer guided tours in an office environment [Hor1993, Hor1996]. Polly's interaction mechanisms were more primitive: users would indicate their will to go on a tour by waving their feet, and the robot would then move around using canned utterances to describe various landmarks. TJ [Tor1994] offered similar functionality, but it could also obey simple stand-alone commands (e.g. "go to the conference room" or "go left") and answer questions about its whereabouts, both typed on a keyboard.

MAIA, a robot that could carry objects from one place to another [ACCF1993, ACC+1994], was also able to obey simple spoken command phrases. Around the same time, the second of the authors was involved in the design of a mobile office assistant at the Microsoft Research Institute of Macquarie University, which could deliver parcels, guide visitors to offices, or offer guided tours. The robot used a commercial dictation system for speech recognition, a state-based dialogue manager, and a language analyzer based on the language interface of [Andr1996]. Jijo-2 [AMF+1999, FAM1998, MAM+1999] is a mobile office assistant with similar capabilities, which can convey information and guide people through an office environment. It communicates in Japanese using a frame-based SDS.

A finer example of robotic assistants is the AESOP 3000 surgical robot [Ver1998]. This is a voice-controlled robot used for delicate work in heart surgery, in effect replacing the hand of the surgeon who controls it by voice. Although it does not provide a full SDS, it is, nonetheless, a very promising example of the growing use of robots in novel environments. Multi-modal interfaces, comprising both speech, keyboard and point-and-click input, have also been employed in recent robots [LBGP2001, PSA+2001].

Studying the robots above led us to the following observations: First, the state of the art in natural language interaction allows usable SDSs to be developed for robots, that advance beyond simple stand-alone commands. Having said this, a second observation is that even without language capabilities, mobile robots can be very complex, involving several subsystems (e.g. navigation, vision, planning) that need to communicate efficiently at real time. This calls for language interaction techniques that are easy to specify and maintain, and that lead to robust and fast language processing. Third, the tasks that most mobile assistants are expected to perform typically require only a limited amount of information from the users; this also applies to Hygeiorobot. These points argue in favor of simple dialogue management approaches, namely state- or frame-based techniques, rather than more complex, plan recognition mechanisms. A fourth observation is that robotic assistants often have to operate in noisy environments (e.g. offices, or in our case, hospital corridors), where they need to interact with many casual users; in our case, hospital staff. This calls for speaker-independent speech recognition and, again, robust language processing.

Overall, then, state- or frame-based approaches appear more suited to mobile robotic assistants like Hygeiorobot's. We have opted for a state-based approach, on the grounds that: (a) a rich development platform for state-based dialogues, which includes a robust speaker-independent recognizer and text to speech synthesis, namely the CSLU Toolkit [SCdV+1998], is freely available for research purposes, and (b) the development time for state-based systems is typically shorter, and Hygeiorobot's SDS had to be constructed in a relatively short time. We also decided not to consider multi-modal input, since our robotic platform does not provide on-board keyboards or pointing-devices.

4 Talking to Hygeiorobot

We now describe in more detail Hygeiorobot's SDS. The SDS allows users to specify all the information that is necessary to deliver a medicine or message to a specific room or patient. The users can also ask for information about the patients, such as the phone or room number of a patient. Furthermore, it is possible to have several tasks pending; for example, it is possible to assign a new delivery to the robot while it is on its way to another room. Figure 4 provides an example of a dialogue with Hygeiorobot.

[H] – Hello, this is Hygeiorobot. Please state your action and priority.
[U] – Medicine delivery at high priority please.
[H] – Please select the room number for the delivery
[U] – 911.
[H] – You have selected room number "911". Is this correct?
[U] – Yes.
[H] – Would you like to give the recipient name?
[U] – Yes. DeForest Kelley.
[H] – So you want me to deliver to room "911" to the person by the name of "DeForest Kelley".
[U] – Yes.

Figure 4: A dialogue with Hygeiorobot

The SDS was developed using the CSLU Toolkit (Figure 5), and it currently supports dialogues in English. Additional speech resources are being developed to support Greek as well. A special mini-grammar was developed for each state, which uses knowledge about the utterances that are likely to be used at each state to identify and extract relevant pieces of information from the user's input. Knowing which words to expect at each state also helps the speech recognizer, which leads to acceptable recognition even in noisy environments.

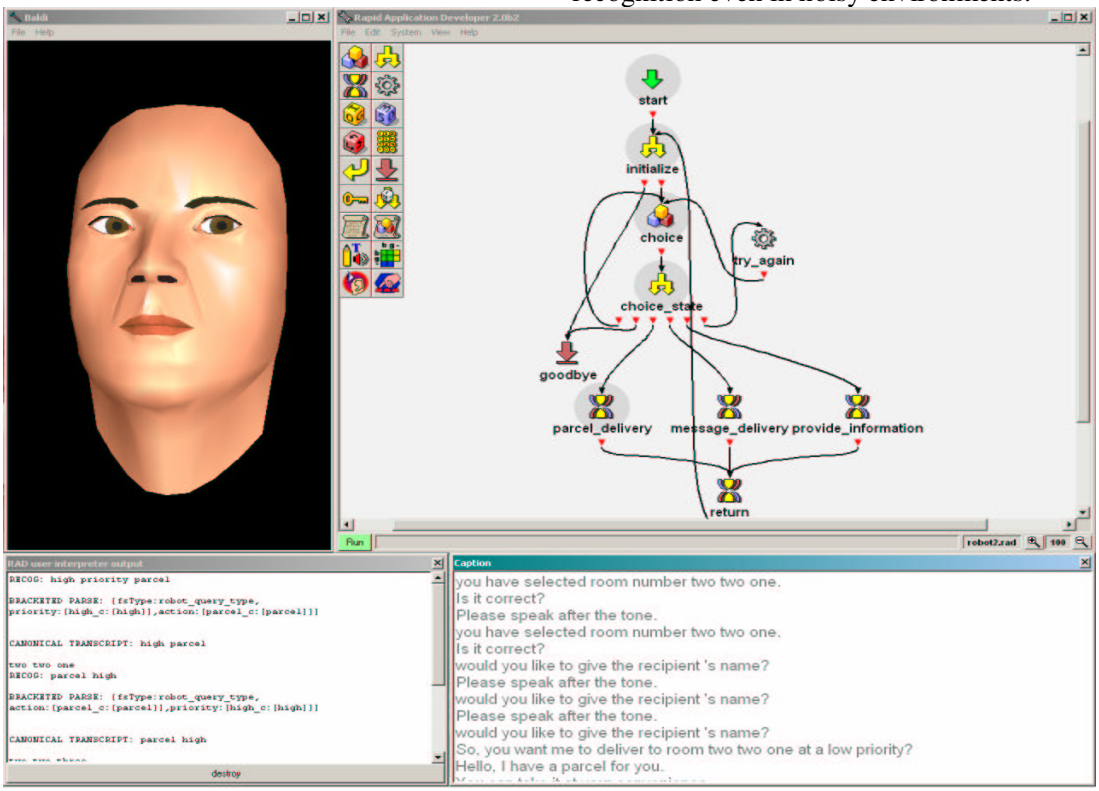

Figure 5: The CSLU-based SDS at work

The dialogue manager was designed to perform relatively short dialogues. The main goal, however, was to ensure that the user's input is interpreted correctly. There are confirmation sub-dialogues at key points to allow the users to check the robot's interpretation of their utterances, and to repeat them if necessary. Additional help and clarification messages are also available.

The SDS has so far been connected to a simulator of the robot's main controller, shown in Figure 6, which allows one to simulate the controller's decisions on which pending task is to be performed, the times when the robot reaches a particular room, etc. Informal tests indicate that the speed and overall performance of the SDS is satisfactory. Integration and field tests with the actual robot are expected to start soon.

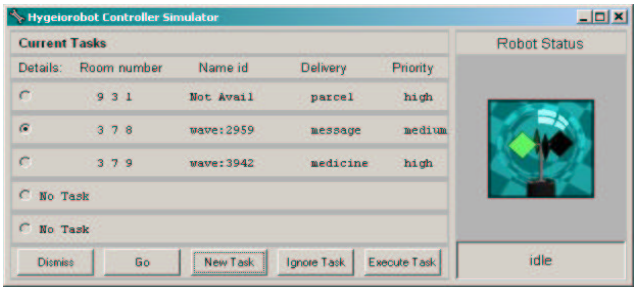

Figure 6: The simulator of the robot's controller

5 Conclusions

We have provided a brief overview of spoken dialogue systems in the context of human-robot interaction. We have also presented a spoken dialogue system intended to allow hospital staff to interact with a robotic assistant that provides information and delivers medicines and messages to

particular rooms or patients. The system has so far been tested with a simulator of the robot's controller. Future project work will be devoted to the integration with the actual robot and field tests. In longer-term work we plan to investigate the usage of machine learning techniques to acquire dialogue management models from corpora of transcribed dialogues.

6 References

[ABD+2001] Allen, J., Byron, D., Dzikovska, M., Ferguson, G., Galescu, L., and A. Stent, (2001). "Towards Conversational Human-Computer Interaction," *AI Magazine*, 2001.

[ACC+1994] Antoniol, G., Caprile, B., Cimatti, A., Fiutem, R., and G. Lazzari, (1994). *Experiencing real-life interaction with the experimental platform of MAIA*. In Proceedings of the 1st European Workshop on Human Comfort and Security, 1994. Held in conjunction with EITC'94.

[ACCF1993] Antoniol, G., Cattoni, R., Cettolo, B., and M. Federico (1993). *Robust Speech Understanding for Robot Telecontrol*. In Proceedings of the 6th International Conference on Advanced Robotics, pages 205–209, Tokyo, Japan, November 1993.

[AMF+1999] Asoh, H., Matsui, T., Fry, J., Asano, F., and S. Hayamizu, (1999) "*A spoken dialog system for a mobile office robot*," Proc. of Eurospeech'99, pp.1139-1142, Budapest, September, 1999.

[AMRS1996] Allen, J.F., Miller, B., Ringger, E., and T. Sikorski, , (1996) "*Robust Understanding in a Dialogue System*," In Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics, 1996.

[And1996] Androutsopoulos, I., (1996). *A Principled Framework for Constructing Natural Language Interfaces to Temporal Databases.* PhD thesis, Department of Artificial Intelligence, University of Edinburgh, 1996.

[AO1995] Aust, H. & M. Oerder, (1995) Dialogue control in automatic inquiry systems. In P. Dalsgaard, L. Larsen, L. Boves & I. Thomsen (eds.*), Proceedings of the ESCA Workshop on Spoken Dialogue Systems*, Vigso, Denmark, 121-124.

[ASF+1994] Allen, J.F., Schubert, L.K., Ferguson, G.M., Heeman, P.A., Hwang, C.H., Kato, T., Light, M.N., Martin, N.G., Miller, B.W., Poesio, M., and D.R. Traum, , (1994), *The TRAINS project: A case study in building a conversational planning agent*. Technical Report 532, Department of Computer Science, University of Rochester, Rochester, NY 14627-0226, September 1994

[BCF+1998] Burgard, W., Cremers, A.B., Fox, D., Hahnel, D., Lakemeyer, G., Schulz, D., Steiner, W., and S. Thrun, (1998) *The interactive museum tour-guide robot.* In Proceedings of the Fifteenth National Conference on Artificial Intelligence, Madison, Wi, 1998.

[FA1993] Ferguson G., and Allen, J.F., (1993) "Generic Plan Recognition for Dialogue Systems," *ARPA Workshop on Human Language Technology*, Princeton, NJ, 21-23 March, 1993

[FAM1998] Fry, J., Asoh, H., and T. Matsui, (1998), *Natural dialogue with the Jijo-2 office robot*, Proceedings of the IROS'98.

[MAM+1999] Matsui, T., Asoh, H., Fry, J., Motomura, Y., Asano, F., Kurita, T., Hara, I., and N. Otsu, (1999) *Integrated natural spoken dialogue system of Jijo-2 mobile robot for office services*. In Proceedings of the AAAI-99 1999.

[Hor1993] Horswill, I., (1993) *Polly: A vision-based artificial agent*. In The Proceedings of the Eleventh National Conference on Artificial Intelligence, 1993.

[Hor1996] Horswill, I., (1996) *The design of the Polly system*. Technical report, Northwestern University, September 1996

[HStD+1996] Hulstijn, J., Steetskamp, R., ter Doest, H., van de Burgt, S., and A. Nijholt, (1996). *Topics in SCHISMA dialogues*. In Proceedings of the Twente Workshop on Language Technology: Dialogue Management in Natural Language Systems (TWLT 11), p. 89-99, 1996

[LBGP2001] Lemon, O., Bracy, A., Gruenstein, A., and S. Peters (2001) *A Multi-Modal Dialogue System for Human-Robot Conversation*, Demo, NAACL2001, June 2001, Pittsburgh, USA

[McT1997] McTear, M. (1997). *Spoken Dialogue Technology: Enabling the Conversational User Interface*. Distributed at the DUS/ELSNET Bullet Course on Designing and Testing Spoken Dialogue Systems, April 1997

[McT1998] McTear, M. (1998). *Modelling spoken dialogues with state transition diagrams: experiences of the CSLU toolkit*. In Proceedings of the International Conference on Spoken Language Processing, volume 4, pages 1223--1226, Sydney, Australia. Australian Speech Science and Technology Association, Incorporated.

[PSA+2001] Perzanowski D., Schultz A., Adams W., Wauchope K., Marsh E. and M. Bugajska (2001) *Interbot: A Multi-Modal Interface to Mobile Robots*, Demo, NAACL2001, June 2001, Pittsburgh, USA

[SCdV+1998] Sutton, S., Cole, R.A., de Villiers, J., Schalkwyk, J., Vermeulen, P., Macon, M., Yan, Y., Kaiser, E., Rundle, B., Shobaki, K., Hosom, J.P., Kain, A., Wouters, J., Massaro, D., and Cohen, M., "*Universal Speech Tools: The CSLU Toolkit*," In Proceedings of the International Conference on Spoken Language Processing (ICSLP-98), vol. 7, pp. 3221-3224, Sydney, Australia, November 1998.

[Tor1994] Torrance, M.C. (1994). *Natural Communication with Mobile Robots,* Master's thesis, Massachusetts Institute of Technology, Cambridge, MA

[Ver1998] Versweyveld, L., (1998) Voice-controlled surgical robot ready to assist in minimally invasive heart surgery, Virtual Medical Worlds Monthly, March 1998

[VvZ1996] Veldhuijzen van Zanten, G. (1996). *Pragmatic interpretation and dialogue management in spoken-language systems*. In Luperfoy, Nijholt, and Veldhuijzen van Zanten, editors, Dialogue Management in Natural Language Systems, TWLT11. University of Twente, p.81-88.