Proceedings of the ACL-02 Demonstrations Session, Philadelphia,
July 2002, pp. 106-107. Association for Computational Linguistics.

Multi-tasking in Practical Multi-modal Dialogue Systems - demonstration system -

Oliver Lemon, Alexander Gruenstein, Alexis Battle, Elizabeth Bratt, and Stanley Peters
lemon,alexgru,ajbattle,ebratt,peters@csli.stanford.edu
CSLI, 210 Panama Street, Stanford University, CA 94305
http://www-csli.stanford.edu/semlab/witas/

1 Introduction

We will demonstrate practical dialogue management techniques for dialogues involving multiple concurrent tasks or activities[1]. Conversational context for concurrent activities is computed using a "Dialogue Move Tree" and an "Activity Tree" which represent multiple interleaved threads of dialogue about different activities and their execution status. Dialogue "threading" also allows the dynamic use of multiple recognition language models, depending on dialogue context – resulting in faster, more efficient recognition. We shall also demonstrate the incremental message selection, aggregation, and generation methods employed in this context (see Figure 1).

The domain of this demonstration is conversational interaction with a robot helicopter, or UAV ('Unmanned Aerial Vehicle') (Doherty et al., 2000). The same dialogue management system is also being used for intelligent tutoring applications, and "in-car" dialogues. This type of application domain is more complex and demanding than the usual information-seeking applications deployed commercially (e.g. ATIS). In particular, interactions with such a system are not scriptable in advance, rely on mixed-initiative in conversation, and may be about multiple interleaved tasks. In such 'practical' dialogues (Allen et al., 2001) we wish to communicate with devices about their possible actions, their plans, goals, and the tasks they are currently attempting. For these reasons we built a dialogue manager that represents (possibly collaborative) activities and their execution status, and tracks multiple threads of dialogue about concurrent and planned activities. This layer of abstraction to "activity models" also allows us to construct a domain-general dialogue move engine, which uses application-specific activity models.

[1]This research was (partially) funded under the Wallenberg laboratory for research on Information Technology and Autonomous Systems (WITAS) Project, Linköping University, by the Wallenberg Foundation, Sweden.

2 Components and Features of the Dialogue System

We shall demonstrate a hub-based architecture (OAA 2.1) with the following software agents:

- *NL (natural language):* a wrapper to SRI's Gemini parser and generator using a grammar for human-robot conversation developed at CSLI.
- *SR (speech recognizer):* a wrapper to a Nuance 8 speech recognition server using a language model compiled directly from the Gemini grammar (with the consequences that every recognized utterance has a logical form, and that every logical form can be mapped to a surface string).
- *TTS (text-to-speech):* a wrapper to the Festival 1.4.3 speech synthesiser, for robot speech output.
- *GUI:* an interactive map display of the current operating environment which displays route plans, waypoints, locations of vehicles including the robot, and allows deictic reference (i.e. mouse pointing) by the user.
- *DM (dialogue manager):* co-ordinates multi-modal inputs from the user, interprets dialogue moves made by the user and robot, updates and maintains the dialogue context, handles robot reports and questions, and sends speech and graphical outputs to the user.
- *Activity Layer:* translates commands and queries from the dialogue interface into commands and queries to the robot, and vice-versa for reports and queries received from the robot. Uses an Activity Model and a realtime CORBA communication layer.

Figure 1 shows some of the system's dialogue capabilities. Our current demonstration system has the following features:

Figure 1: A demonstration of the WITAS dialogue system (2002 version)

Multi-modal Utterances	**Dialogue Moves**
User (U): *Our job is to look for a red car*	Raise Joint Activity
Robot (R): *Ok. I am looking for one.*	Confirm Activity
U: *Fly here please [+click on map]*	Deictic command
R: *Okay. I will fly to waypoint one*	Confirm Activity, introduce referent
R: *Now taking off and flying there.*	Report Multi-tasking Activity
U: *Stop that. Go to the tower instead.*	Command, Revision
R: *I have cancelled flying to waypoint one. I will fly to the tower.*	Report Activity status.
U: *What are you doing?*	Activity query
R: *I am searching for a red car and flying to the tower*	Answer
U: *What will you do next?*	Activity query
R: *I have nothing planned.*	Answer
R: *I see a red car on main street* [display on map, show video images], *Is this the right car?*	Report, Joint-Activity query
U: *Yes, that's the right car*	Confirm complete activity
R: *Okay. I am following it .*	Report initiated activity

- supports multi-tasking, multiple topics, and collaboration,
- support of commands, questions, revisions, and reports, over a dynamic environment,
- multi-modal, mixed-initiative, open-ended dialogues,
- Semantic-Head-Driven Generation of system reports
- echoic and variable message generation, filtered for relevance and recency
- asynchronous, real-time operation.

Details of the system architecture can be found in (Lemon et al., 2001a; Lemon et al., 2001b), the Dialogue Move Tree and Activity Models are explained in (Lemon et al., 2002 submitted), and possible applications in interactive entertainment are discussed in (Lemon, 2002).

Further details concerning the demonstration version of the system (including video) are available at: `http://www-csli.stanford.edu/semlab/witas/`.

3 Demo requirements

A monitor and speakers for display purposes: the demonstration system resides on a notebook PC under Windows 2000, and requires no special hardware or connectivity.

References

James Allen, Donna Byron, Myroslva Dzikovska, George Ferguson, Lucian Galescu, and Amanda Stent. 2001. Toward conversational human-computer interaction. *AI Magazine*, 22(4):27–37.

Patrick Doherty, Gösta Granlund, Krzystof Kuchcinski, Erik Sandewall, Klas Nordberg, Erik Skarman, and Johan Wiklund. 2000. The WITAS unmanned aerial vehicle project. In *European Conference on Artificial Intelligence (ECAI 2000)*.

Oliver Lemon, Anne Bracy, Alexander Gruenstein, and Stanley Peters. 2001a. Information states in a multi-modal dialogue system for human-robot conversation. In Peter Kühnlein, Hans Reiser, and Henk Zeevat, editors, *5th Workshop on Formal Semantics and Pragmatics of Dialogue (Bi-Dialog 2001)*, pages 57 – 67.

Oliver Lemon, Anne Bracy, Alexander Gruenstein, and Stanley Peters. 2001b. The WITAS Multi-Modal Dialogue System I. In *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech' 01)*, Aalborg.

Oliver Lemon, Alexander Gruenstein, and Stanley Peters. 2002 (submitted). Collaborative actvities and multi-tasking in dialogue systems. *Traitement Automatique des Langues (TAL)*. Special Issue on Dialogue.

Oliver Lemon. 2002. Transferable multi-modal dialogue systems for interactive entertainment. In *AAAI Spring Symposium on Artificial Intelligence in Interactive Entertainment*, Technical Report SS-02-01, pages 57 – 61, Menlo Park, CA. AAAI Press.