

A Flexible Conversational Dialog System for MP3 Player

Fuliang Weng¹ Lawrence Cavedon² Badri Raghunathan¹ Danilo Mirkovic² Ben Bei¹
Heather Pon-Barry¹ Harry Bratt³ Hua Cheng² Hauke Schmidt¹ Rohit Mishra⁴ Brian Lathrop⁴
Qi Zhang¹ Tobias Scheideck¹ Kui Xu¹ Tess Hand-Bender¹ Sandra Upson¹ Stanley Peters²
Liz Shriberg³ Carsten Bergmann⁴

Research and Technology Center, Robert Bosch Corp., Palo Alto, California¹
Center for Study of Language and Information, Stanford University, Stanford, California²
Speech Technology and Research Lab, SRI International, Menlo Park, California³
Electronics Research Lab, Volkswagen of America, Palo Alto, California⁴

{Fuliang.weng,badri.raghunathan,hauke.Schmidt}@rtc.bosch.com
{lcavedon,huac,peters}@csli.Stanford.edu
{harry,ees}@speech.sri.com
{rohit.mishra,carsten.bergmann}@vw.com

1 Abstract

In recent years, an increasing number of new devices have found their way into the cars we drive. Speech-operated devices in particular provide a great service to drivers by minimizing distraction, so that they can keep their hands on the wheel and their eyes on the road. This presentation will demonstrate our latest development of an in-car dialog system for an MP3 player designed under a joint research effort from Bosch RTC, VW ERL, Stanford CSLI, and SRI STAR Lab funded by NIST ATP [Weng et al 2004] with this goal in mind. This project has developed a number of new technologies, some of which are already incorporated in the system. These include: end-pointing with prosodic cues, error identification and recovering strategies, flexible multi-threaded, multi-device dialog management, and content optimization and organization strategies. A number of important language phenomena are also covered in the system's functionality. For instance, one may use words relying on context, such as 'this,' 'that,' 'it,' and 'them,' to reference items mentioned in particular use contexts. Different types of verbal revision are also permitted by the system, providing a great convenience to its users. The system supports multi-threaded dialogs so that users can diverge to a different topic before the current one is finished and still come back to the first after the second topic is done. To lower the cognitive load on the

drivers, the content optimization component organizes any information given to users based on ontological structures, and may also refine users' queries via various strategies. Domain knowledge is represented using OWL, a web ontology language recommended by W3C, which should greatly facilitate its portability to new domains.

The spoken dialog system consists of a number of components (see Fig. 1 for details). Instead of the hub architecture employed by Communicator projects [Senef et al, 1998], it is developed in Java and uses a flexible event-based, message-oriented middleware. This allows for dynamic registration of new components. Among the component modules in Figure 1, we use the Nuance speech recognition engine with class-based ngrams and dynamic grammars, and the Nuance Vocalizer as the TTS engine. The Speech Enhancer removes noises and echo. The Prosody module will provide additional features to the Natural Language Understanding (NLU) and Dialogue Manager (DM) modules to improve their performance.

The NLU module takes a sequence of recognized words and tags, performs a deep linguistic analysis with probabilistic models, and produces an XML-based semantic feature structure representation. Parallel to the deep analysis, a topic classifier assigns top n topics to the utterance, which are used in the cases where the dialog manager cannot make

any sense of the parsed structure. The NLU module also supports dynamic updates of the knowledge base.

The CSLI DM module mediates and manages interaction. It uses the dialogue-move approach to maintain dialogue context, which is then used to interpret incoming utterances (including fragments and revisions), resolve NPs, construct salient responses, track issues, etc. Dialogue states can also be used to bias SR expectation and improve SR performance, as has been performed in previous applications of the DM. Detailed descriptions of the DM can be found in [Lemon et al 2002; Mirkovic & Cavedon 2005].

The Knowledge Manager (KM) controls access to knowledge base sources (such as domain knowledge and device information) and their updates. Domain knowledge is structured according to domain-dependent ontologies. The current KM makes use of OWL, a W3C standard, to represent the ontological relationships between domain entities. Protégé (<http://protege.stanford.edu>), a domain-independent ontology tool, is used to maintain the ontology offline. In a typical interaction, the DM converts a user's query into a semantic frame (i.e. a set of semantic constraints) and sends this to the KM via the content optimizer.

The Content Optimization module acts as an intermediary between the dialogue management

module and the knowledge management module during the query process. It receives semantic frames from the DM, resolves possible ambiguities, and queries the KM. Depending on the items in the query result as well as the configurable properties, the module selects and performs an appropriate optimization strategy.

Early evaluation shows that the system has a task completion rate of 80% on 11 tasks of MP3 player domain, ranging from playing requests to music database queries. Porting to a restaurant selection domain is currently under way.

References

- Seneff, Stephanie, Ed Hurley, Raymond Lau, Christine Pao, Philipp Schmid, and Victor Zue, *GALAXY-II: A Reference Architecture for Conversational System Development*, International Conference on Spoken Language Processing (ICSLP), Sydney, Australia, December 1998.
- Lemon, Oliver, Alex Gruenstein, and Stanley Peters, *Collaborative activities and multi-tasking in dialogue systems*, Traitement Automatique des Langues (TAL), 43(2), 2002.
- Mirkovic, Danilo, and Lawrence Cavedon, *Practical Multi-Domain, Multi-Device Dialogue Management*, Submitted for publication, April 2005.
- Weng, Fuliang, Lawrence Cavedon, Badri Raghunathan, Hua Cheng, Hauke Schmidt, Danilo Mirkovic, et al., *Developing a conversational dialogue system for cognitively overloaded users*, International Conference on Spoken Language Processing (ICSLP), Jeju, Korea, October 2004.

