The Pragmatics of Taking a Spoken Language System Out of the Laboratory

Jody J. Daniels and Helen Wright Hastie
Lockheed Martin Advanced Technology Laboratories
3 Executive Campus, 6th Floor
Cherry Hill, NJ 08002
{jdaniels, hhastie}@atl.lmco.com

Abstract

Lockheed Martin’s Advanced Technology Laboratories has been designing, developing, testing, and evaluating spoken language understanding systems in several unique operational environments over the past five years. Through these experiences we have encountered numerous challenges in making each system become an integral part of a user’s operations. In this paper, we discuss these challenges and report how we overcame them with respect to a number of domains.

1 Introduction

Lockheed Martin’s Advanced Technology Laboratories (ATL) has been designing, developing, testing, and evaluating spoken language understanding systems (SLS) in several unique operational environments over the past five years. This model of human interaction is referred to as Listen, Communicate, Show (LCS). In an LCS system, the computer listens for information requests, communicates with the user and networked information resources to compute user-centered solutions, and shows tailored visualizations to individual users. Through developing these systems, we have encountered numerous challenges in making each system become an integral part of a user’s operations. For example, Figure 1 shows the deployment of a dialogue system for placing Marine supply requests that is being used in a tactical HMMWV.

One of the challenges of creating such spoken language systems includes giving appropriate system responses. This involves managing the tension between utterance brevity (for rapid completion of tasks or if on a secure channel) and giving enough context in a response to build the user’s trust (to prove that the system really did understand and correctly interpret the user’s utterance). Similarly, users don’t want to have to repeat redundant information, which would add to the electronic signature of the soldier. The system must be robust when handling Out Of Vocabulary (OOV) terms and concepts. It must also be able to adapt to noisy environments whose parameters change frequently and be able use input devices and power access unique to each situation.

2 Architecture

The LCS Spoken Language systems use the Galaxy architecture (Seneff et al., 1999). This Galaxy architecture consists of a central hub and servers. Each of the servers performs a specific function, such as converting audio speech into a text translation of that speech or combining the user’s past statements with what was said most recently. The individual servers exchange information by sending messages through the hub. These messages contain information to be sent to other servers as well as information used to determine what server or servers should be contacted next.

Various Galaxy Servers work together to develop a semantic understanding of the user’s statements and questions. The sound spoken into the microphone, telephone, or radio is collected by an Audio Server and sent on to the recognizer. The recognizer translates this wave file into text that is sent to a natural language parser. The parser converts the text into a semantic frame, a representation of the statement’s meaning. This meaning representation is passed on to the Dialogue Manager that monitors the current context of a conversation and, based on this context, can prompt the user for any necessary clarification and present intelligent responses to the user. Since the Dialogue Manager is aware of what information has been...
discussed thus far, it is able to determine what information is still needed. A semantic frame is created by the Dialogue Manager. This frame is sent through the Language Generation Server to generate a text response which is then spoken to the user through a speech synthesis server.

To solve the problem of retrieving or placing data from/in remote and local sources, we gave the systems below the use of mobile software agents (Daniels, 2000). If user-requested information is not immediately available, an agent can monitor the data sources until it is possible to respond. Users may request a notification or alert when a particular activity occurs, which may happen at an indeterminate time in the future. Because of the potentially significant time lag, it is important to manage dialogue activity so that the user is only interrupted when the need for information is more important than the current task that the user is currently undertaking. This active management of interruptions aids task management and lightens cognitive load (Daniels et al., 2002).

3 Domains

3.1 LCS Marine

One of the first LCS systems to be tested in the field was our Marine Logistics spoken dialogue system. This application sought to connect the Marine in the field to the Small Unit Logistics (SUL) database, which maintains current information about supply requisitions. Warfighters wanted to be able to place requests as well as check on the status of existing requests without the need of additional hardware or communicating with a third party. It was also highly desirable to use existing communications procedures, so that the training time to use the system was minimized. The system needed to be speaker-independent and mixed initiative, enabling the warfighters to develop a sense of trust in the technology.

Marines using the system were able to perform several tasks. They could create new requests for supplies, with the system prompting them for information needed to fill in a request form. They could also modify and delete previously placed requests and could check on the status of requests in one of two ways. They could directly ask about the current status, or they could delegate an agent to monitor the status of a particular request. It was easy to build the constraint of having the agent return after a specified time period if no activity occurs on the request, which is also valuable information for the Marine. These delegated agents travel across a low-bandwidth SINCGARS radio network from the Marine to the database and access that database to place, alter, and monitor supply requisitions.

The challenges in deploying this system to the field were twofold - building trust in the system so that it would become part of normal operations and in dealing with the unique environmental factors. The former presented the conflicting goals of brevity versus confirming user inputs. Marines want to restrict their time on the radio net as much as possible. At the same time they want to ensure that what they requested is what they were going to receive. Much time went into defining and refining system responses that met both needs as best possible. This involved several sessions with numerous Marines evaluating possible dialogue responses. We also spent much time ensuring that LCS Marine could handle both proper and non-standard radio protocols. Broad coverage of potential expressions, especially those when under stress, such as recognition of the liberal use of curse words, led to greater user ability to successfully interact through the system.

The second set of challenges, unique environmental factors, included access while on the move, battlefield noise, and coping with adverse conditions such as sand storms. Accessing LCS Marine while on the move meant using a SINCGARS radio as the input device. Attempts to use the system by directly collecting speech from a SINCGARS radio were dropped due to the technological challenges presented by the distortion introduced by the radio on the signal. Instead, we installed the majority of the system on laptops and put these into the HMMWV. We sent mobile agents over the SINCGARS data link back to the data sources. This meant securing hardware in a HMMWV and powering it off of the vehicle’s battery as illustrated in Figure 1. (Only one laptop was damaged during testing.) The mobile agents were able to easily traverse a retransmission link and reach the remote data source.

Dealing with hugely varying background noise sources was less of a problem than originally predicted. Fortunately, most of the time that one of these loud events would occur, users would simply stop talking. Their hearing was impaired and so they would wait for the noise to abate and then continue the dialogue. On the other hand, we did encounter several users who, because of the Lombard effect, insisted upon always yelling at the system. While we did not measure decibel levels, there were a few times when the system was not able to understand the user because of background noise.

3.2 Shipboard Information

An LCS system has also been developed to monitor shipboard system information aboard the Sea Shadow (IX 529), a Naval test platform for stealth, automation, and control technologies. From anywhere on the ship, personnel use the on-board intercom to contact this system, SUSIE (Shipboard Ubiquitous Speech Interface Environment), to ask about the status of equipment that is located throughout the ship. Crew members do not have to be anywhere near the equipment being monitored in order to receive data. Figure 2 illustrates a sailor using SUSIE through the ship’s intercom.

Personnel can ask about pressures, temperatures, and voltages of various pieces of equipment or can delegate monitoring those measurements (sensor readings) to the system. A user can request notification of an abnormal reading by a sensor. This causes the LCS system to delegate a persistent agent to monitor the sensor and to report the requested data. Should an abnormal reading occur, the user is contacted by the system, again using the intercom.

Personnel can ask about pressures, temperatures, and voltages of various pieces of equipment or can delegate monitoring those measurements (sensor readings) to the system. A user can request notification of an abnormal reading by a sensor. This causes the LCS system to delegate a persistent agent to monitor the sensor and to report the requested data. Should an abnormal reading occur, the user is contacted by the system, again using the intercom.

This domain presented several challenges and opportunities. Through numerous discussions with users and presentation of possible dialogues, we learned that the users would benefit from a system ability to remember, between sessions, the most recent activity of each user. This would permit a user to simply log in and request: “What about now?” SUSIE would determine what had been this user’s most recent monitoring activity, would then seek out the current status, and then report it. While
this seems quite simple, there is significant behind-the-scenes work to store context and make the interaction appear seamless.

It was necessary to use the organic intercom system installed in the Sea Shadow for communication with crew members. Collecting speech data through the intercom system to pass to SUSIE required linking two DSP (digital signal processors) (and adjusting them) to the hardware for the SLS. Once connected, the next significant challenge was that of the varying noise levels. Background noise varied from one room to the next and even within a single space. We were not able to use a push-to-talk or hold-to-talk paradigm because of the inconvenience to the crew members; they leave the intercom open for the duration of the conversation. Fortunately, the recognizer (built on MIT’s SUMMIT (Glass and Hazen, 1998)) is able to handle a great deal of a noise and still hypothesize accurately. To improve the system accuracy, we will incorporate automatic retraining of the recognizer on noise each time that a new session begins.

3.3 Battlefield Casualty Reporting System

We are currently developing a new LCS system known as the Battlefield Casualty Reporting System or BCRS. The goal of this system is to assist military personnel in reporting battlefield casualties directly into a main database. This involves intelligent dialogue to reduce ambiguity, resolve constraints, and refine searches on individual names and the circumstances surrounding the casualty. Prior knowledge of every individual’s name will not be possible. The deployment of this system will again present many challenges such as noise effects on a battlefield, effects of stress on the voice, and the ability to integrate into existing military hardware.

4 Future Work

Areas of research needed for more dynamic and robust systems include better, more robust or partial parsing mechanisms. This would allow systems to ask the user to clarify just that portion of an utterance which was either misunderstood or not included in the lexicon, rather than discarding the entire utterance. Systems must be able to cope with multi-sentence inputs, such as humans do. Most current SLS only permit one utterance at a time to be given by a user. With longer utterances, the system should provide backchannels to confirm that the system is still listening.

Ease of domain expansion continues to be important as systems evolve. Requiring that system modifications be done solely by developers will inhibit the expansion, adaptation, and utility of a SLS over time. Additionally, the ability to work in variable operational environments means that deployable systems will need additional noise adaptation or mitigation techniques. Further, the ability to switch modes of communication if one is not appropriate at a given time will increase a system’s operational value.

5 Conclusions

We have discussed the pragmatics involved with taking an SLS system out of the laboratory or away from telephony and placing it in a volatile environment. These systems have to be robust and be able to cope with varying input styles and modes as well as be able to modify their output to the appropriate situation. In addition, the systems must be an integral part of the technology that is in current use and be able to withstand adverse conditions. Satisfying all of these constraints involves active participation in the development process with the end-users as well as creative solutions and technological advances.

6 Acknowledgments

These systems were developed and tested by a number of excellent engineers, including Steve Knott, Dan Miksch, Mike Orr, Jose Rivera, Kathy Stibler, and Mike Thomas. This work was supported by DARPA contract N66001-01-D-6011.

References

Jody Daniels, Susan Regli, and Jerry Franke. 2002. Support for intelligent interruption and augmented context recovery. In Proceedings of 7th IEEE Conference on Human Factors and Power Plants, Scottsdale, AZ, September.

Jody Daniels. 2000. Integrating a spoken language system with agents for operational information access. In Proceedings of Innovative Applications of Artificial Intelligence (IAAI-2000), Austin, TX, August.

Jim Glass and T.J. Hazen. 1998. Telephone-based conversational speech recognition in the jupiter domain. In Proceedings of ICSLP 98, Sydney, Australia, November.

Stephanie Seneff, Ray Lau, and Joe Polifroni. 1999. Organization, communication, and control in the galaxy-ii conversational system. In Proceedings for Eurospeech ’98, Budapest, Hungary.