IAC-04-T.3.07

CLARISSA SPOKEN DIALOGUE SYSTEM FOR PROCEDURE READING AND NAVIGATION

James L. Hieronymus
NASA Ames Research Center, USA
jimh@email.arc.nasa.gov

John Dowding
UARC University of California, Santa Cruz, USA
jdowding@email.arc.nasa.gov

MS T27A-2, NASA Ames Research Center
Moffett Field, CA 95035

ABSTRACT

Speech is the most natural modality for humans use to communicate with other people, agents and complex systems. A spoken dialogue system must be robust to noise and able to mimic human conversational behavior, like correcting misunderstandings, answering simple questions about the task and understanding most well formed inquiries or commands. The system aims to understand the meaning of the human utterance, and if it does not, then it discards the utterance as being meant for someone else. The first operational system is Clarissa, a conversational procedure reader and navigator, which will be used in a System Development Test Objective (SDTO) on the International Space Station (ISS) during Expedition 10. In the present environment one astronaut reads the procedure on a Manual Procedure Viewer (MPV) or paper, and has to stop to read or turn pages, shifting focus from the task. Clarissa is designed to read and navigate ISS procedures entirely with speech, while the astronaut has his eyes and hands engaged in performing the task. The system also provides an MPV like graphical interface so the procedure can be read visually. A demo of the system will be given.

FULL TEXT

Introduction

Future exploration missions will require that spacecraft and planetary surface operations become self sustaining, with minimal assistance from the ground. This will be necessary to control costs of long duration missions and because of large communication time delays of up to 40 minutes for the Mars mission. Conversational intelligent agents offer a way to multiply astronaut effectiveness by allowing tasks and monitoring to be delegated to the agents. The agents could report the ongoing status and progress of operations toward achieving mission goals.

The most natural way to interact with these intelligent agents is by conversing with them, sometimes augmented with displays, keyboards and touch screens. There are some activities which are best commanded using speech and others by pointing to locations on displays or pointing to a step in an operation. Often it is quicker to read a lengthy instruction, note, caution or warning from a display, than to have the system read it. Numbers seem to be more easily remembered when seen rather than listened to.

A spoken dialogue system takes spoken input from the human user, interprets the meaning of the utterance in the context of the task being done, and then presents the results of the operation to the user. Such systems which have been developed with DARPA funding include Air Travel planning, restaurant and weather information, and air transport scheduling. These research systems have not included conversational behavior necessary for always successfully accomplishing the task.

Familiar commercial systems for obtaining flight arrival and departure information, obtaining stock quotes and making stock trades, and finding telephone numbers have been available for about 5 years. These systems tend to be fragile, and many users do not get the information they request or are transferred to a human operator. Most of the flight arrival and departure timetable systems work best if you have the flight number. This sort of fragile behavior is clearly unacceptable for space applications.

Spoken dialogue systems for space should mimic human conversational behavior, so as to lessen the training required. The usual methods for repairing misunderstandings should work with the system. For example, if the system misrecognizes an utterance, the user can say "I meant Step 19," and the system should undo whatever it did and perform the same operation on step 19. The speech recognition grammars should allow any sensible utterance to be understood and acted upon. NASA acronyms and abbreviations are included in the system using approved lists and by recording astronauts reading procedures to each other.

The system should be robust to the type of noise found on the International Space Station (ISS). Active noise canceling microphones are sufficient to assure high quality speech recognition and Clarissa has been tested in ISS recorded noise accurate to 1 dB in each octave band. The system performs well in both the Service Module and the US Lab Module, with only a few percent increase in word error rate. Because utterances are recognized for their meaning, this slightly lower performance has a small effect on usability.

Procedures are the central way in which operations are done on the ISS. Each task to do with maintenance, testing water and air quality, repairing a system on the ISS, or checking the EVA Suits (EMU's) is written up as a detailed procedure. A typical procedure might have 100 steps, 20 branch points, and some values to be entered into the system. Presently the procedures are written in word and are also available as pdf files. For the International Procedure Viewer (IPV) the procedures are being converted to XML which allows different levels of description.

On the ISS procedures are currently done by one astronaut who reads the procedure from a laptop computer procedure reader (Manual Procedure Viewer, (MPV) currently, and

International Procedure Viewer (IPV) soon) or from a printed procedure manual. This requires that he take his eyes off of the equipment being used and find the place in the procedure, read the next step and then do it. For often done procedures the text only serves to remind the astronaut of what to do next. At the bottom of a page the crew member must press a page down key or un-Velcro the manual, turn the page and Velcro it back down. For someone using a glove box, this requires depressurizing the gloves which can take up to a minute.

Clarissa Procedure Assistant

The Clarissa conversational astronaut assistant[1,2] is meant to talk astronauts through important procedures and checklists for maintenance, repair and monitoring of ISS systems. Currently it can read procedures one step at a time, navigate to arbitrary steps and sub steps, ask for branching decisions and proceed down the selected branch, read ahead in the procedure while maintaining its place, correct misrecognitions when the user says "I meant go to step 7", state the present place in the procedure when asked "Where was I?," start a new procedure while putting the present procedure in the background, record and play voice notes, set timers, stop talking when asked to "Shut Up" and control the volume of the output when requested to "Speak Up." There is a Challenge Verify Mode for crucial steps in the procedure, which tracks the completion of each step and will not allow the astronaut to proceed without saying that the step has been completed, and Terse Mode, which reads only step titles for someone very experienced in a particular procedure. There is also a facility for showing pictures and diagrams which are accessed with a "Show me the" command. There are help commands which tell the legal commands at this point in the dialogue, and detailed help which explains the present options in detail. Clarissa can take voice notes to

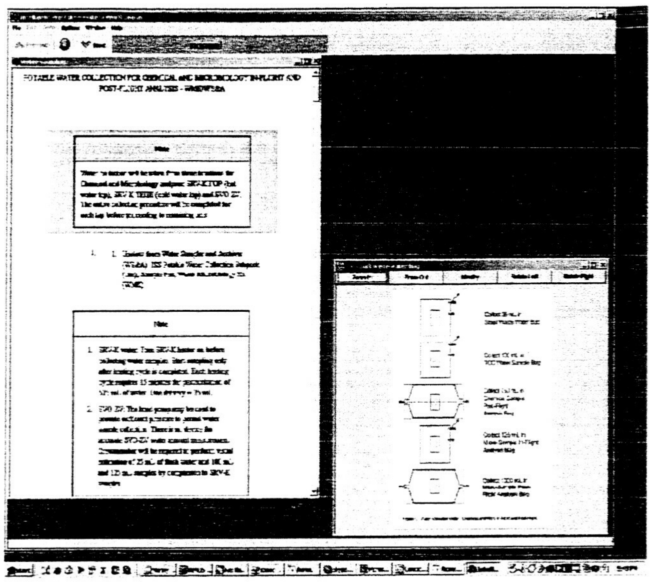

Figure 1: Clarissa Screen for Procedure

detail changes in a procedure or at training time for a personal note on how best to do a procedure. These notes will be available on orbit. Clarissa is scheduled to be an SDTO on Expedition 10 with 5 procedures, WMK Nominal Water processing, WMK Inflight Coliform Detection, WMAK/WMK Visual Analysis, EMU Checkout and LCVG Water Fill.

Figure 1 shows a Clarissa screen for one of the Water Analysis procedures. The current step is highlighted in green on the left panel, and the diagram on the lower right shows the water bags associated with the procedure.

In order to have a usable system several components are necessary, the ability respond only to commands intended for the system while ignoring the rest which is called open microphone speech dialogue, the ability to keep track of context so that normal indirect references can be made , navigating the procedure by step number and sub-step number with the step read and displayed, ability to answer help questions and the ability to accept and test values which the crew members put into the system.

Open Microphone Spoken Dialogues

One requirement for conversational spoken dialogues is that the system should know when you are speaking to it and when you are speaking to someone else. The Clarissa system determines whether on not an utterance has meaning for it in the present context and if it does not, discards the utterance as speech intended for someone else. This means that the system does not try to respond to every utterance, and that conversations with other crew members can happen without system interruption. In tests where conversational speech was played to the system, it was rejected 96% of the time. Further work is reducing the false accept rate even further, by taking into account more of the context. For example if a yes-no question has not been asked, then a yes-no response is not for the system, and must be part of another conversation. The use of support vector machine techniques developed by Jean-Michel Render[3] is also helping to reduce the false accept rate. The speech recognition engine used in Clarissa is from Nuance Communications[4].

Clarissa also has the suspend command to make it only respond to the resume command. This is useful if there is a long conversation with another crew member about some other topic. At the conclusion of the discussion the system can be turned back on for the task being done.

Implementing Clarissa Procedures

Presently the written procedure needs to be processed by hand in order to make an XML representation of the procedure. This includes the step execution structure and the spoken part of the procedure. The XML version must preserve the step text exactly, the step and sup-step numbering exactly, and provide what is said by the system.

Written procedures are difficult to understand when spoken, since they are written to be read visually. In order to make them understandable auditorily it is sometimes necessary to paraphrase them. For example if there is a branch point and the written procedure gives "If the suit delta P is greater than .3, perform leak repair procedure." Clarissa asks "Is the suit delta P greater that .3?" A "yes" response results in Clarissa saying go to the leak repair procedure.

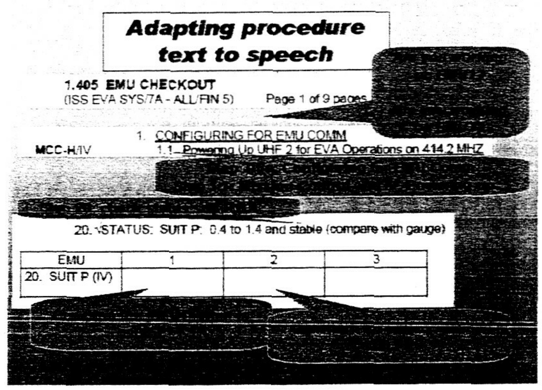

Figure 2: EMU Checkout

An example of this is shown in Figure 2 for asking about which EMU are being checked and then subsequently asking questions only about the suits being checked.

These paraphrases have to be checked by the procedure writers to versify that they are true to the original meaning of the procedure text. Then the resulting XML procedure is placed under version control by the SODF.

After the procedure is implemented in Clarissa it has to be Procedure Verified by the procedure writers and the crew office. This provides more feedback about how understandable the paraphrase is and assures that the resulting procedure is done correctly using Clarissa.

Finally the paraphrased text of the procedure has to be verified as being what is supposed to be said. This is done having humans listen to the utterance and vote on whether the text is what is said. Secondly by using speech recognition to verify that the correct work string occurred in the read utterance.

This two pronged approach takes care of two problems, the tendency for listeners to correct speech errors unconsciously and for speech recognizers to give bad match scores if the reader's pronunciation is not the standard one used. On the one hand if two people have understood the read speech to be the text string desired, then it is likely that the read utterance will be correctly understood. If the speech recognition shows that all of the correct words are in the recognized utterance with the correct order, it is also more likely that the read utterance is correct. By cross checking both of these methods we are assured that the procedure steps are read accurately.

Noise Tests

A large auditorium at NASA Ames was filled with recorded ISS noise from three locations, the Service Module, the US Lab and the US Airlock. The power levels were accurate to 1 dB in each octave from 120 – 16000 Hz. Thirty seven subjects used the Clarissa system to do two of the procedures, one in Service Module noise and one in US Lab noise. The second experiment was to test the subject's ability to understand the spoken procedures by repeating back what they had heard in Airlock noise. The subjects wore Bose Aviation X headsets with active noise canceling microphones. People tend to speak at a volume level which is approximately 20 dB above the background noise, a phenomena called the Lombard effect. Here the problem is that with the noise canceling headset, the environment sounds quiet, so the person speaks quieter, but the noise canceling microphone is in the loud noise. The result was that the system was useable at all noise levels which represented the locations. Service module noise was louder, at 74 dB than the US Lab module at 65 dB. At both levels the noise which was recorded on the ISS consisted of a combination of fan noise and mechanical noise including pumps and motors. Listening to the recorded wave files, the noise was apparently very soft in the background.

Foreign Accents

Approximately 16 astronauts participated in trials of the Clarissa system. In informal trials Japanese and

Spanish astronauts were able to use the system with an acceptably low error rate. Tests with Russian students showed degraded performance for Russian accented English. The word error rate was perhaps twice the native speaker error rate, and lead to many corrections. Usually foreign accented talkers find expressions which the system recognizes and stick with those. Eventually a Russian accented English recognizer will be developed to allow high quality use of Clarissa by Cosmonauts.

Conclusion

The Clarissa procedure navigator and reader has been developed for a SDTO on ISS Expedition 10. The system has been developed with feedback and suggestions from many astronauts. The present system has been shown to be robust to ISS environmental noise, and can support an astronaut in carrying out a complex procedure.

Acknowledgements:

The Clarissa Team:
Manny Rayner, Beth Ann Hockey, Nikos Chatzichrisafis, Kim Farrell and the authors

Astronauts: TJ Creamer, Mike Fincke, Tim Kopra, Jim Voss, Steve Lindsey, Wendy Lawrence, Neil Woodward, Janice Voss, Stan Love, Megan McArthur, Janet Kavandi, Jim Reilly, Andrew Feustel, Nick Patrick, Jeff Ashby, Don Pettit, Peggy Whitson, Shannon Lucid, Nicole Stott, William McArthur and Leroy Chiao

Training: Rick Davis, Mary Baxter, David Chiquelin, Derone Jenkins-Todd, Michael Root

SPOC: Frank Wood, Robert Readiger
CRCP: Sean Kelly, Mark Swaby, Saera Quershi, Christine Vo
Engineering: Lui Wang, Jeff Graham, Dave Kortenkamp
Medical OPS: Jack Smith, Kathy Johnson, Carissa Vaughn, Cindy Woo
EVA: Roger Lottridge and Mark Lipsky

NASA Ames - JSC liaison: Kevin Bass

Xerox Research Center Europe:
Jean-Michel Render

Student Interns:
Amy Fischer, Vladimir Tkachenko, Claire Castillo

1. Rayner, M., Hockey, B.A., Hieronymus, J., Dowding, J. and Aist, G., "An Intelligent Procedure Assistant Built using REGULUS 2 and ALTERF, Proc. ACL, Sapporo, Japan, July 2003.
2. Dowding, J. and Hieronymus, J., "A Spoken Dialogue Interface to a Geologist's Field Assistant," Proc. HLT/NAACL, Edmonton, AL. May 2003.
3. Jean-Michel Render, "Kernel Methods in Natural Language Processing," ACL-04 Tutorial, Barcelona, Spain, July 2004.