Proceedings of HLT-NAACL 2003
Demonstrations , pp. 9-10
Edmonton, May-June 2003

A Spoken Dialogue Interface to a Geologist's Field Assistant

John Dowding and James Hieronymus
Mail Stop T-27A-2
NASA Ames Research Center
Moffett Field, CA 94035-1000
{jdowding,jimh}@riacs.edu

Abstract

We will demonstrate a spoken dialogue interface to a Geologist's Field Assistant that is being developed as part of NASA's Mobile Agents project. The assistant consists of a robot and an agent system which helps an astronaut wearing a planetary space suit while conducting a geological exploration. The primary technical challanges relating to spoken dialogue systems that arise in this project are speech recognition in noise, open-microphone, and recording voice annotations. This system is capable of discriminating between speech intended for the system and for other purposes.

Start tracking my g p s coordinates.
Start logging my bio sensors every fifteen seconds.
Where is my my current location?
Call this location Asparagus.
Create a new sample bag and label it sample bag three.
Take a voice note
Please begin recording voice note now:
This sample originated in a dry creek bed. [pause]
Would you like to continue recording the voice note?
no
Voice note has been created.
Associate that voice note with sample bag three.
Play the voice note associated with sample bag three.

Table 1: Example utterances

1 Introduction

The Geologist's Field Assistant is one component of Mobile Agents, a NASA project studying technologies, techniques, and work practices for sophisticated human-agent and human-robot cooperation in space and planetary exploration environments, such as the surface of Mars. The evolution, development and evaluation of this project occurs in a series of increasingly complex field tests in Mars analog environments (deserts and artic sites) on Earth.

The Spoken Dialogue component assists an astronaut wearing a space suit while conducting a geological exploration, by tagging samples by spoken discriptions, commanding the taking of pictures, recording descriptive voice annotations, and tracking the associations between these samples, images, and annotations. The assistant will also help track the astronaut's location and progress though the survey, and help track their body exertion level (heart and respiration rate).

The Spoken Dialogue interface is one of several of Mobile Agents, each with different goals and evaluation metrics. Other components include: Brahms work-practice modelling and simulation system (NASA Ames); MEX mobile wireless networking (Ames); robots (Johnson Space Center (JSC) and Georgia Tech); Spacesuits (JSC) and Biomedical sensors (Stanford University); Satellite Internet services (Goddard Space Flight Center); and Geologists (US Geological Survey).

The primary technical challanges relating to spoken dialogue systems that arise in this project are open-microphone speech recognition and understanding which decides which agent receives, and responds to a particular utterance and space suit noise.

2 Example Dialogue

The language capabilities developed so far are largely direct commanding with the user controlling task initiative. A sample of user commands is given in Table 1. A system response is always given, but is usually omitted below for the sake of brevity. When given, the system response appears in italics.

3 Architecture

This spoken dialogue system shares a common architecture with several prior systems: CommandTalk (Stent et al., 1999), PSA (Rayner et al., 2000), WITAS (Lemon et al., 2001), and the Intelligent Procedure Assistant (Aist et al., 2002). The architecure has been well described in

prior work. The critical feature of the architecture relevant to this work is the use of a grammar-based language model for speech recognition that is automatically derived from the same Unification Grammar that is used for parsing and interpretation.

4 Data Collection

The Mobile Agents project conducted two field tests in 2002: a one week dress rehearsal at JSC in the Mars yard in May, and a two week field test in the Arizona desert in September, split between two sites of geological interest, one near the Petrified Forest National Park, and the other on the ejecta field at Meteor Crater. We collected approxmimately 5,000 recorded sound files from 8 subjects during the September tests, some from space-suit subjects, and the rest in shirt-sleeve walk-throughs (still a high wind condition). We transcribed 1059 wave files. All conditions were performed open-mic and all sounds that were picked up by the microphone were recorded, so not all of these files contained within-domain utterances intended. Of the transcribed sound files, 208 contained no speech (mostly wind noise) and 243 contained out-of-domain speech that was intended for other hearers. That left 608 within-domain utterances that were split 80%-20% into test and training utterances.

5 Technical Challanges

The Geologist's Field Assitant requires the ability to make voice notes that can be stored and transmitted. We implemented this by adding a recording mode to the speech recognizer agent, and temporarily increasing the speech end-pointing interval. This allows us to record multi-sentence voice notes without treating inter-sentence pauses as end-of-voice-note markers. Entering recording mode is triggered by specific speach acts like *Take a voice note* or *Annotate sample bag one.*

When considering recognition accuracy in the open-mic condition, we consider additional metrics beyond word-error rate (WER). Since the recognizer can fail to find a hypothesis for an utterance, we compute the false-rejection rate (FREJ) for within-domain utterances and adjusted word-error (AWER) counting only the word errors on the non-rejected utterances. We also consider misrecognitions of out-of-domain utterance as within-domain, and compute the false-accept rate (FACC). Table 2 gives the performance results for the grammar-based language model that was used in the September test. This model gives reasonable performance on within-domain utterances, but falsely accepts 25.5% of out-of-domain utterances. After the September test, we used the training data we had collected to build a Probabilistic Context-Free Grammar using the `compute-grammar-probs` tool that comes with Nuance (Nuance, 2002). Using only 485 utterances of training data, there was improvement in both the AWER and FACC rates, resulting in a language model where both FREF and FACC were under 10%. There was also a substantial improvement in recognition speed, as measured in multiples of CPU real-time.

Version	WER (%)	FREJ (%)	AWER (%)	FACC (%)	xCPUrt (%)
	Baseline CFG Language Model				
Training	12.56	4.54	7.72	—	58.9
Test	9.5	3.25	7.5	25.5	57.5
	Probabilistic CFG Language Model				
Training	9.97	5.57	4.6	—	19.4
Test	8.99	7.32	3.7	9.09	19.0

Table 2: Comparing Baseline and Probabilistic CFG

6 Conclusions

We will demonstrate a dialogue system that has an improved ability to discriminate between speech that is intended for different purposes, treating some as data objects to be saved, and others identified as being out-of-domain. With probabilities on the rules the system has an acceptably low false accept rate and is fast and accurate.

References

G. Aist, J. Dowding, B.A. Hockey, and J. Hieronymus. 2002. An intelligent procedure assistant for astronaut training and support. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (demo track)*, Philadelphia, PA.

O. Lemon, A. Bracy, A. Gruenstein, and S. Peters. 2001. Multimodal dialogues with intelligent agents in dynamic environments: the WITAS conversational interface. In *Proceedings of 2nd Meeting of the North American Association for Computational Linguistics*, Pittsburgh, PA.

Nuance, 2002. http://www.nuance.com. As of 15 November 2002.

M. Rayner, B.A. Hockey, and F. James. 2000. A compact architecture for dialogue management based on scripts and meta-outputs. In *Proceedings of the 6th Applied Natural Language Processing Conference*, Seattle, WA.

A. Stent, J. Dowding, J. Gawron, E. Bratt, and R. Moore. 1999. The CommandTalk spoken dialogue system. In *Proceedings of the Thirty-Seventh Annual Meeting of the Association for Computational Linguistics*, pages 183–190.