"Excuse me, where's the registration desk?" Report on Integrating Systems for the Robot Challenge AAAI 2002

**D. Perzanowski, A. Schultz, W. Adams, M. Bugajska,
M. Abramson,
M. MacMahon,
A. Atrash,
M. Coblenz**

Naval Research Laboratory, Washington DC

<dennisp | schultz | adams | magda | abramson> @aic.nrl.navy.mil
adastra@mail.utexas.edu
amin@cc.gatech.edu
coblenz@andrew.cmu.edu

Abstract

In July and August 2002, five research groups—Carnegie Mellon University, Northwestern University, Swarthmore College, Metrica, Inc., and the Naval Research Laboratory—collaborated and integrated their various robotic systems and interfaces to attempt The Robot Challenge held at the AAAI 2002 annual conference in Edmonton, Alberta. The goal of this year's Robot Challenge was to have a robot dropped off at the conference site entrance; negotiate its way at the site, using queries and interactions with humans and visual cues from signs, to the conference registration area; register for the conference; and then give a talk. Issues regarding human/robot interaction and interfaces, navigation, mobility, vision, to name but a few relevant technologies to achieve such a task, were put to the test.

In this report we, the team from the Naval Research Laboratory, will focus on our portion of The Robot Challenge. We will discuss some lessons learned from collaborating and integrating our system with our research collaborators, as well as discuss what actually transpired—what worked and what failed—during the robot's interactions with conference attendees in achieving goals. We will also discuss some of the informal findings and observations collected at the conference during the interaction and navigation of the robot to complete its various goals.

Introduction

Edmonton, situated on the westernmost portion of the Prairie provinces and capital of the Province of Alberta, Canada and boasting the world's largest indoor shopping mall, hosted the Eighteenth National Conference on Artificial Intelligence. As part of each year's AAAI activities at the conference, a Robot Challenge is sponsored. The purpose of the Challenge is to push the state-of-the-art in robotics research. The purpose of this year's event, chaired by Ben Kuipers of the University of Texas at Austin, was to get a robot to register for the conference and to give a talk about itself at the event.

> … a robot will start at the entrance to the conference center, need to find the registration desk, register for the conference, perform volunteer duties as required, then report at a prescribed time in a conference hall to give a talk.
>
> --*AAAI-2002 Mobile Robot Competition & Exhibition* flyer

These ambitious goals clearly would push the technology, since as everyone in robotics research currently knows, getting a robot to do one thing successfully is a major achievement, let alone a host of inter-related goals. Attempting this year's Robot Challenge meant that whoever approached the problem had a robot or was building one that could tackle such a large host of tasks.

Did such a robot exist or was one waiting in the wings?

Well, sort of a little of both as far as we were concerned.

In this report, we will give an overview of the entire task, as it was set out for The Robot Challenge and present a brief overview of the tasks as they were apportioned (more or less consensually) to each of the participating research institutions in our group. Next, we will present in more detail the particular portion of the task for which our particular sub-group assumed responsibility. We will then discuss our successes and failures in the goals of our part of The Robot Challenge (the former being by far the

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2002		**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
"Excuse me, where's the registration desk?" Report on Integrating Systems for the Robot Challenge AAAI 2002	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,4555 Overlook Ave SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

In July and August 2002, five research groups-Carnegie Mellon University, Northwestern University, Swarthmore College, Metrica, Inc., and the Naval Research Laboratory-collaborated and integrated their various robotic systems and interfaces to attempt The Robot Challenge held at the AAAI 2002 annual conference in Edmonton, Alberta. The goal of this year's Robot Challenge was to have a robot dropped off at the conference site entrance; negotiate its way at the site, using queries and interactions with humans and visual cues from signs, to the conference registration area; register for the conference; and then give a talk. Issues regarding human/robot interaction and interfaces, navigation, mobility, vision, to name but a few relevant technologies to achieve such a task, were put to the test. In this report we, the team from the Naval Research Laboratory, will focus on our portion of The Robot Challenge. We will discuss some lessons learned from collaborating and integrating our system with our research collaborators, as well as discuss what actually transpired-what worked and what failed-during the robot's interactions with conference attendees in achieving goals. We will also discuss some of the informal findings and observations collected at the conference during the interaction and navigation of the robot to complete its various goals.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**10**	

shortest section of this report). Finally, we will present some of the anecdotal lessons learned from our participation in this event.

The Groups Organize

One of the groups to take on this year's Robot Challenge was a motley crew consisting of AI researchers, and graduate and undergraduate students from several institutions. Led by the indefatigable and undaunted Reid Simmons (Carnegie Mellon University), Ian Horswill (Northwestern University), Bruce Maxwell (Swarthmore College), Bryn Wolfe (TRACLabs), and Alan Schultz (Naval Research Laboratory), a collaboration of research groups was formed.

It was their belief that an integrated robotic system could be constructed to accomplish

all [timestamp: early April 2002]…

most [timestamp: late May 2002]…

many [timestamp: mid-June 2002]…

some [timestamp: 3 a.m. on July 31, 2002] of the goals set forth in The Robot Challenge.

So what's so hard about finding a registration desk?

In a brief aside now, image the first time you ever registered for a conference. You may have been in a foreign or unknown city. If anything, the venue for the conference was probably an unknown site. You probably floundered around a bit trying to find your way. You may have looked for some signs or asked an official-looking individual at the center or a conference-attendee-type where the registration desk was. If no satisfactory response was obtained or you politely nodded that you understood their mumbled or convoluted instructions and continued to bumble around the center, you eventually found your way and were able to register for this first most important event.

But look at what was involved in just these first few steps in trying to find your way around an unknown environment. These are some of the underlying capabilities that our robot must exhibit, if it is to do anything in this event.

Inside the venue, you looked for a sign. But in robotic terms that involves machine vision. Wow! Bottleneck number one. As a human, you are equipped with an incredibly complex but efficient means for obtaining visual. You have eyes and can look for signs, and then once you find a sign, you can read it.

Well, let's gloss over this vision problem, for the moment, and let's say that instead of trying to look for and read a sign, let's say you looked for an official-looking person at the venue. What does an "official-looking" person look like? The person is wearing a uniform, let's say. Pushing aside all of the problems of how a robot "finds" or senses a human, let's assume you know what a person is and you don't go walking up to a pillar to ask for directions, you look for a human wearing a uniform. Think of how much knowledge is necessary to encode the concept of a uniform, which might be a standard uniform for that site but is probably not a universal instance of what a "uniform" is. Oh, the official at the site has a badge. Well, now we're back at the vision problem with having to find and read signs. Let's not do that loop again. Instead, let's rely on the kindness of strangers and just walk up to a human and ask for directions.

If the person is wearing a conference badge, there's a good chance that they know where the registration desk is. How else did they get their badge? Have someone else get it for them? Steal one? But there's that ugly vision problem again—finding a sign, in this case a badge and being able to read it.

So let's just use tact here and come right out and ask anything that resembles a human where the registration desk is. And you get a response to your query!

However, did you have problems understanding the human? Chances are you did. Especially if that person doesn't speak the same native language or dialect as you. Another bottleneck for robotics emerges: the speech recognition problem. But with repeated responses for clarification or repetition where necessary, you do eventually get the verbal instructions for the directions to the registration desk. Next comes the task of navigating to the goal.

Robotic navigation comes with its own set of problems: sensors, localization, map adaptation, local versus global navigation, etc. However, these problems are tractable and you can manage to find your way to the registration area.

But now you have to get in the right line. That means you have to read a sign, or ask. If your vision capabilities are strong, you can read and get into the appropriate line, or you can ask someone to point you in the right direction. As a human, it's pretty easy for you to know where the end of a line is. As a human, you have already been socially conditioned not to just barge right in to the front of a line, but you find your way to the end of it and take your place, perhaps interacting pleasantly with your nearest neighbor as you wait your turn. After all, you do want to network. Humans are pretty social and sociable creatures. We do seem to organize in groups and we do congregate for events and share refreshments and chit-chat. When given the opportunity, they seem to like to converse with each other about one thing or another. In the parlance, this is called "schmoozing," and some humans are better at it than others, but robots have no concept of schmoozing whatsoever. Given that schmoozing can focus on any topic, it is quite a feat to get a robot to converse along one topic of conversation and then jump to something else, retrace conversational steps or go boldly off into a schmoozing realm where no robot has ever gone before. Given a very limited capability in this area currently, most robots will stand politely silent or mouth Eliza-like platitudes just to keep a human's attention, waiting for the next environmental input that prompts the firing of an already programmed algorithm for action.

Moving up in line as fellow attendees peel off the queue, you can eventually interact with the person behind the

desk, registering for the conference and politely asking where to go to give your talk.

Again you must rely upon your navigational abilities to find the location for your talk, but then once there you are back on familiar territory—you have already prepared your slides, you know your material, and you can deliver your presentation, hopefully without any interruptions from the audience. If you're lucky, the only time you have to interact with another human now, is during the question-and-answer period at the end of your presentation. However, if the audience asks pointed questions about specifics in your talk, you already have the knowledge to answer and can provide an intelligent response.

The Birth of GRACE

In April 2002, our team convened for the first time and began tackling the problem. Somewhere along the line (off-line or on-line) a name emerged for the robot entity they were creating, GRACE (Graduate Robot Attending ConferencE) [Figure 1].

Figure 1. Amazing GRACE

GRACE, "a B21R Mobile Robot built by RW, has an expressive face on a panning platform as well as a large array of sensors. The sensors include a microphone, touch sensors, infrared sensors, sonar sensors, a scanning laser range finder, a stereo camera head on a pan-tilt unit, and a single color camera with pan-tilt-zoom capability. GRACE can speak using a high-quality speech synthesizer, and understand responses using her microphone and speech recognition software"[1].

The group analyzed the task of registering for a conference and giving a talk. They basically saw it as a five-part process. (A similar breakdown and participation of the event is offered at our website summarizing the event. See Note 1.) The first part was getting the directions, and navigating to the conference registration area. (Getting directions and navigating to the room for presentation was considered another incarnation of this task). The second part was navigating to the desk itself. The third part was negotiating the line for registration. The fourth was the actual act of registering (this could actually be considered similar to the first part, but we separated it for expositional purposes). Finally, the fifth part was presenting a talk at the conference. As each part of the entire process of registering for the conference and giving the talk were analyzed, we noticed that various member research institutions could tackle certain parts of the larger task more appropriately than others. Each institution had its own research strengths, and utilizing these strengths became the keystone for building the final integrated system to accomplish The Robot Challenge.

Thus, the Carnegie Mellon University (CMU) research team concentrated their efforts on navigational skills. Based on their work in navigation, CMU handled basic navigation and the ability to detect and negotiate riding an elevator at the conference center. CMU also assumed the research responsibility of getting GRACE to stand in line (Nakauchi and Simmons 2002), converse with the registrar and register, navigate to the assigned location for her talk, and express herself to humans via an onscreen image of her face (Bruce, Nourbakhsh, and Simmons 2002).

For GRACE's ability to communicate with humans, the Naval Research Laboratory (NRL) relied on its work in human-robot interaction (Perzanowski et al. 2002; Perzanowski, Schultz, and Adams 1998) and provided the human-robot interface that permitted spoken natural language and natural gestures. With this interface, a human could give the robot directions verbally, as well as point in the right direction for any of the locations desired. Furthermore, the interface allowed for clarifications and corroborations of information. Thus, for example, if an utterance was unclear, GRACE could indicate this by saying "What?" or when presented with a particular goal, GRACE could ask "Am I at the registration desk?" Furthermore, if the human told GRACE to turn left, for example, but then pointed to the right, GRACE could complain, "You told me to turn left but pointed right. What do you want me to do?"

GRACE was also outfitted with a Personal Digital Assistant (PDA). NRL has been using a wireless PDA along with its natural language interface to provide gestures [Figure 2].

Figure 2. Wireless Personal Digital Assistant with stylus

Based on the information obtained from the environment, NRL's PDA interface builds up a localized map which is displayed as a touch screen on the PDA. With it, users can provide gestural input: a set of x,y coordinates on a map with an utterance such as "Go over here" or "Follow this path" accompanied by a line drawn on the map, or "Go to this doorway" accompanied by a stylus tapping on an x,y location on the map. Natural gestures are sensed by high resolution rangefinders and were to be seen by stereo cameras, but these capabilities were not available in time for GRACE's performance.

So, our PDA interface was revised to sweeping gestures using a stylus on the PDA touch screen to indicate left and right directions. Unfortunately, due to programmer errors before the event, the device was rendered inoperable for the Challenge. For the record, we did have this device coded up and tested, but our current hypothesis about the failure is that the process talking to it was never started.

For the people-tracking parts of the task and for locating faces, TRACLabs provided a vision sensor and appropriate controllers. However, given the various other sensors and controllers that were already installed on GRACE, these were disabled on site.

Swarthmore's work on vision and recognition using OCR was utilized to locate and interpret signs and badges, as well as to implement visual servoing capabilities used to approach the registration desk. Northwestern provided a software package that generates a verbal presentation with accompanying PowerPoint slides generated by plugging together canned text strings it obtains from a database.

NRL under the Gun

In the remainder of this report, we will concentrate on NRL's portion of The Robot Challenge; namely, getting GRACE from the entrance of the conference center to the registration area.

GRACE was not allowed to have any prior knowledge of the layout of the conference center but was allowed general knowledge about conference centers; for example, they may consist of multiple floors and if so, there are typically three ways to get between them: stairs, escalator, and elevator. With this information we immediately started working on getting GRACE to understand all of the navigation and directional issues involved in getting downstairs via an elevator.

Since GRACE cannot navigate stairs or escalators, our options for getting to the goal were mercifully limited. We knew further that we had to navigate down inside the conference center. All of this basically helped us ensure that our grammars and dictionaries would handle likely human-robot interactions.

GRACE was supposed to find the registration area through interactions with humans. She could have scoured the crowd for an official, but then we had the problem of trying to figure out what an official might look like. So, we rejected that option outright. Next, we considered the possibility of GRACE's scouring the entrance area for a conference badge and then interacting with the badge holder to get directions to the registration area. But at the time, when we were considering this option, some of the later available software was still in beta-testing, so we optioned for a different interaction initially.

GRACE was going to scour the area for a human—she could do that given the sensors and functionalities of the other modules being provided by our co-researchers. She was then going to interact with whomever she saw and get as much information out of this informant as possible. She would follow the instructions given to her, and if she managed to find the area from this one interaction, the goal would be achieved. If more information was needed, she would then go to another human and interact accordingly.

All of this sounded very nice and normal to us. That's basically what people do. However, for verbal interactions with our robots back at NRL, we had been using an off-the-shelf speech recognition product. Granted, it had been honed and fine tuned via a Speech Development Kit, so its recognition rate for its limited vocabulary and grammatical constructions was extremely high. While we have no quantified data to support this claim, our general feeling and the feeling of people using this modality to interact with our robots back in the lab was that the speech recognition device we were using was pretty impressive. Human users felt comfortable interacting with the robots in our lab using speech, and they didn't seem overly frustrated with poor recognition of utterances.

While we felt pretty confident about the speech recognition device we were using, we were still not confident enough to let it ride unbridled at the conference. After all, it had been trained on one researcher's voice: male; exhibiting a standard US English dialect (if there is such an animal). And it worked in the lab with other males exhibiting dialects tolerably close to whatever dialect it was that the trainer provided. However, we feared we might have problems if a female were to be chosen by GRACE to interact with. We could have trained GRACE with a female trainer as well, but then we would have had to provide a switch in the recognition engine which would load the appropriate models. All of this, while do-able, wasn't done. Other matters came up. It was pushed on the stack, and didn't get to pop.

Furthermore, if anyone had a slight accent, the speech recognition engine would have problems. We also worried about paraphrasability. After all, who said "How many ways do I love thee?" We started sweating about the myriad of different ways somebody might tell GRACE to do something. While the prototype grammars and dictionaries that were provided to GRACE had been worked on for many years in our lab, and are therefore rather robust in being able to capture a great deal of the richness of language, we tended to shy on the side of conservatism and decided to have GRACE interact with one of us who was fairly familiar with GRACE's linguistic capabilities, as rich as they might be. We knew she could handle a great deal of paraphrasability: that's one of the strengths of the semantic component of our natural language understanding system, Nautilus (Wauchope 1994; Perzanowski et al. 2001). However, we just didn't want to take the chance of putting it to the test in a very conversational environment with a complete stranger, far beyond what it had ever been exposed to. Furthermore, since this was the initial step of a long registration task with plenty of other places to go wrong and our team members relying on a competent (and successful) first step, we, therefore, opted to have GRACE interact with one of our own team members who was fairly familiar with GRACE's linguistic capabilities and fairly comfortable with the speech hardware in general.

Given the ability to interact verbally, GRACE also needed the ability to detect the presence and location of humans. This information was provided by one of our team members, CMU. However, we had already built in a gesture-detection component into our human-robot interface, and we utilized it here in order to disambiguate any speech that required some kind of gesture to accompany it. For example, if the human said "Grace, it's over there," but didn't point to any location in the real world, GRACE is capable of asking for additional information, or at least complaining that something is amiss by asking "Where?" Likewise, GRACE can detect inconsistencies in verbal and gestural input, as, for example, if a person says "Grace, turn left" but points right, GRACE will complain, "I'm sorry, you told me to turn one way, but pointed in another direction. What do you want me to do?"

These were capabilities that we brought to the drawing board with us. However, what we needed to do for The Challenge was to integrate our capabilities (expand them where necessary) and the capabilities and modules of other systems from our co-researchers' efforts and create a unified system.

For our portion of the task, we created a module written using TDL libraries. Our module interacts with other modules through an IPC interface which was developed specifically for The Challenge. Our task module therefore interacts with the other modules of GRACE's architecture, such as mobility, and object recognition and the facial module displaying GRACE's facial expressions and movements on the computer screen, as in Figure 3.

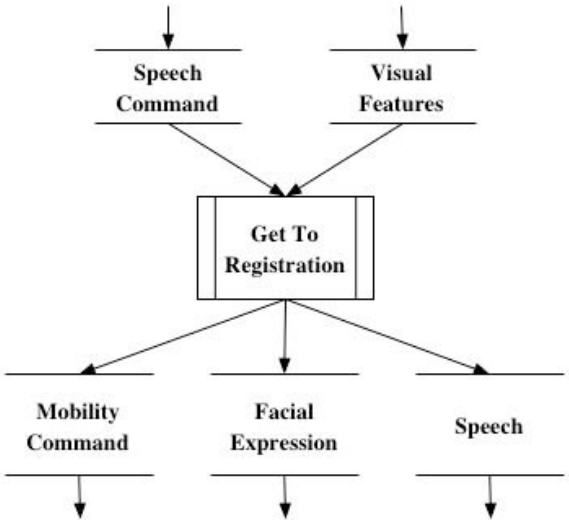

Figure 3. Interaction of various modules

Our task module, Figure 4, interleaves linguistic and visual information with direction execution. Given a specific destination to be reached, such as a registration area, the task module interleaves the information gathering with direction execution.

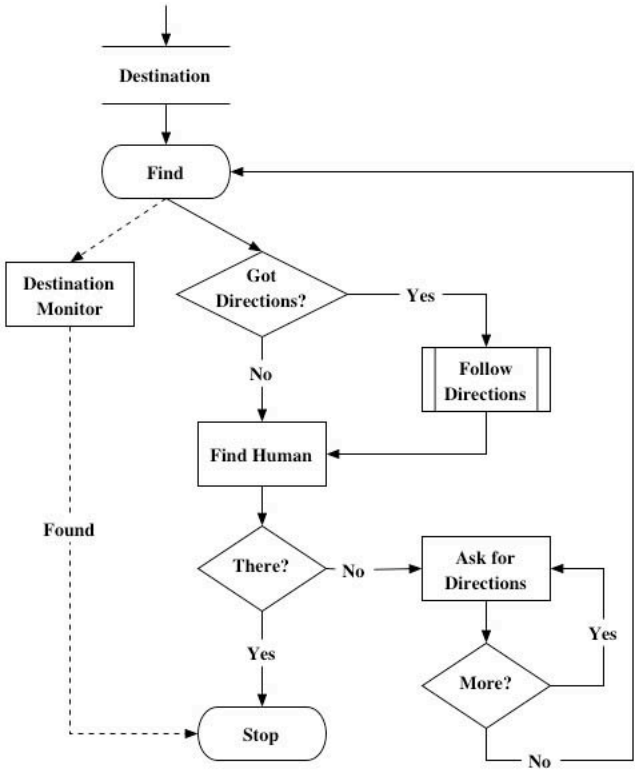

Figure 4. NRL Task Module

The module allows for a priori knowledge in the form of a destination. For GRACE this was the registration desk. If there are no directions to be followed, GRACE performs a random walk until a human is detected. Once a human is found, she starts a conversation, the goal of which is to obtain the directions to the destination in question. The directions can include simple commands such as "Turn left," "Go forward five meters," as well as higher level instructions such as "Take the elevator," or "Turn left next to the Starbucks." Once a list of directions is complete, they are executed on the beginning of next iteration of the control loop. The task is completed once the destination is reached, as determined by an explicit human confirmation or perception of the goal.

Parallel to this loop, there are a number of monitors running; for example, an explicit "STOP" command can be issued if unforeseen or dangerous conditions arise; perception processing occurs, allowing the detection of the destination or a human.

The directions obtained during the information-gathering stage of the task are associated with a specific destination. They are executed sequentially in the order in which they were given by the human.

There are two types of direction that can be given. First, there is a simple action command, such as "Turn left." We assume that in order to get to the destination, GRACE should execute this command before executing the next instruction. The second type of command is an instruction specifying an intermediate destination, such as "Take the elevator to the second floor." In this case, a new intermediate goal is instantiated (the elevator), and the logic is recursively applied to the new goal. Once all the available directions have been executed and successfully completed, GRACE can conclude that she has either arrived at the destination or additional information is required to reach the goal. If GRACE perceives the destination before all the directions are executed, the remaining ones are abandoned, and she continues with the next goal.

Thus, if GRACE asks a human bystander, "Excuse me, where is the registration desk," and the human responds, "Grace, to get to the registration desk, go over there <accompanied by a gesture>, take the elevator to the ground floor, turn right, and go forward fifty meters," in terms of our module as described here, the human's input is mapped to a representation something like the following:

Find Registration Desk:
  Find Elevator (ground floor);
    Go over there <gesture>;
  Turn right;
  Go forward 50 meters.

GRACE in the Spotlight

With our system as outlined in the previous section, we worked on integrating it with the other components provided by our co-workers from the other member institutions of our team. On the day of The Robot Challenge, GRACE took her place at the entrance to the conference site.

The site chosen for GRACE's entrance was two flights above the conference registration level. As shown in Figure 5, GRACE's task would require that she navigate from the upper right-hand portion of the photo to the lower portion of the building. This would require that she find out from one of the bystanders how to get down to the lower level.

The dialog for this part of the task basically was as follows:

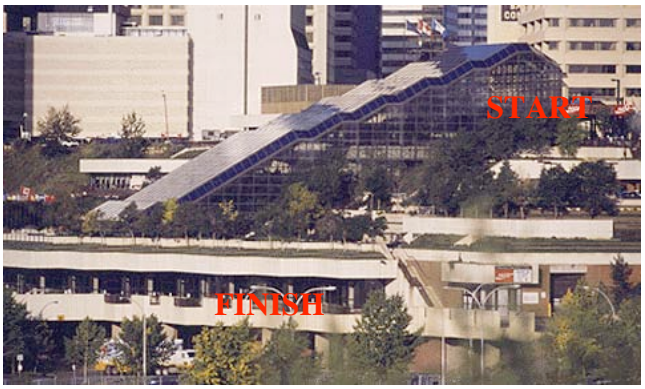

Figure 5. Exterior View of Shaw Convention Centre, Edmonton

GRACE: Excuse me, could you tell me how to get to the registration desk?
Human: Grace, take the elevator down two flights.
GRACE: Where is the elevator?
Human: It's over there <accompanying gesture>.

See Figure 6.

Figure 6. GRACE and Human (on left) interact and negotiate boarding the elevator (out of sight on right)

The above dialog basically characterizes the initial interchange between GRACE and our researcher chosen to interact with her.

Because we knew beforehand that our speech recognition software and hardware would have some difficulty in this environment, we asked the management of the conference center to turn off a rather lovely but noisy waterfall that ran alongside the elevator, escalators and stairs, and cascaded down from the street level where GRACE started to the conference level two floors below, Figure 7.

Figure 7. Top-view inside elevator looking down to registration level with waterfalls to either side; escalator and stairs on right

After the initial dialog, GRACE had some difficulty navigating the passageway onto the elevator, but she succeeded, and asked for manual assistance pressing the correct button down to the appropriate level (GRACE does not have manual capabilities at this time).

GRACE's ability to detect the elevator, both in getting on and off using Carnegie Mellon University's elevator-detecting algorithms was quite impressive, Figure 8.

Figure 8. View of closed elevator doors

Once down on the conference level and off the elevator, GRACE had some problems negotiating her way through a narrow L-shaped passageway from the elevator onto the main concourse level. The human accompanying her at this point tried to use natural language and gestures to navigate through the narrow area. GRACE was basically told to turn left or right or go straight ahead, and gestures accompanied the commands where appropriate. However, because she was having navigation problems and the gestures from the PDA were not working as expected, the human seemed to have become frustrated. This caused the speech software to have difficulty, and GRACE had trouble understanding the human. Tension and stress in the human voice can cause many state-of-the-art speech recognition systems to experience problems. This is precisely what happened at this point in GRACE's journey.

After a few tense moments of misunderstood communication, both GRACE and the human managed to navigate the passageway, and she was able to turn a couple of corners, walk several meters forward and turn to face the registration area and desk, Figure 9.

Figure 9. GRACE (center behind Human) arrives at the registration area

At this point, our task was ended. We had managed to assist GRACE in navigating from an area two flights above, onto an elevator, and down to the registration area. It was now our team member's task to get GRACE to register and then find her pre-assigned room to give her talk.

While our task was over, suffice it to say here that GRACE did indeed register for the conference. She managed to find the appropriate line to stand in, Figure 10.

Figure 10. GRACE rehearses reading "Robots" and registering for conference

She read the correct sign that was for "ROBOTS," using software developed by the Swarthmore team, but perhaps in robotic enthusiasm, she rudely barged into the line of people waiting to be registered, who were actually judges for the event, instead of politely taking her place at the end of the line.

After interacting with the registration booth personnel, GRACE navigated to the conference exhibit hall where she first stopped to visit a vendor's booth, then continued to the bleachers at the rear of the hall where she gave her presentation, using a program designed by Northwestern University, Figure 11.

Figure 11. GRACE (foreground right with back to camera) delivers talk at AAAI 2002. Audience members view large screen above right (out of view) of PowerPoint presentation

She eventually received several awards: The Technology Integration Award, The Human-Computer Interaction Award, and The Award for Robustness in Recovery from Action and Localization Errors. Everyone attending found the latter award rather humorous and perversely appropriate, given GRACE's somewhat rude behavior in the registration line.

The Morning After

From our point of view and the tasks we were responsible for attempting and completing, two basic lessons-learned emerged in hindsight:

1. The hardest part of this task was obtaining unambiguous directions to the registration area. Great care had to be taken in the design of the system to make sure that in conversation with the human, GRACE would ask questions that produced clear, simple directions, but still facilitate fairly robust and flexible conversation.
2. While a general planner to tackle our part of the task was designed, a more robust general planner is desirable, especially one that can handle more cases and one that can provide the robot with more behaviors.

The Gloriously Optimistic Future

Given our experiences with our portion of The Robot Challenge we are currently underway investigating some improvements to GRACE's architecture.

Our research at NRL in natural language and gesture has thus far involved writing robust grammars that can withstand a great deal of paraphrasing. However, it is next to impossible to predict the richness and complexity of human speech and interaction with just a phrase structure grammar and pre-defined lexical items. In a sense, therefore, we have been shooting with one hand tied behind our backs. Our one gun has been loaded with a natural language understanding system that can interpret what is being said, but we also need a natural language system that is robust enough to parse larger elements, namely a dialog, so that the unpredictable nature (as well as the predictable nature) (Grosz and Sidner 1986; Grosz, Hunsberger and Kraus 1999) of human conversation and interaction can be captured. In terms of speech recognition, we await this community's accomplishments in developing robust systems capable of handling speech in noisy environments, as well as with true cross-speaker capabilities. We are currently working with Sphinx (Huang et al. 1993) to create a speech recognition system that does not have to be trained and to develop grammars capable of handling novel utterances. Further, to achieve some of these goals we will interleave some stochastic parsing at the dialog level, since we believe this information can help in disambiguating some utterances and ultimately provide for more robust natural language parsing and full semantic interpretation of utterances.

In terms of our desire to incorporate more general planning into our architecture, we are working on one based more on probabilistic reasoning (Likhachev and Arkin 2001), rather than the more-or-less finite state machine we utilized in this year's Robot Challenge.

Finally, we assume that human interaction is facilitated by each individual in the interaction having a similar or comparable cognitive model of the interaction. Since human behavior can be characterized and modeled, and because we believe this behavior can be used as a model of behavior in robots that interact with humans, we are investigating ways of incorporating a cognitive model for human-robot interaction written in ACT-R (Anderson and Lebiere 1998).

Some Additional Thoughts

The following are some additional thoughts gleaned from comments of our team members about their work and participation in The Robot Challenge.

When asked what the researchers thought was the hardest part of his or her task, one team member said that the building of our architecture for the task was complex because we were dealing with constantly evolving building

blocks.□Furthermore, the TDL language, control-level robot skills, and interfaces and predicted abilities of other software modules were constantly in flux. Another team member said that the hardest part was exploring all possible cases the robot might encounter, and all possible paths through the decision space. Since we used a finite state machine to navigate through the decision space, we need to think about ways to avoid hard-coding a finite state machine and do it a bit more intelligently. Finite state machines get big quickly, and it's very difficult to be sure that the whole thing works. All states really need to be tested, but in a large system this is infeasible and impractical. This line of thinking parallels our earlier conclusion that probabilistic reasoning and cognitive modeling might help us around this issue.

When dealing with a large number of people at different sites, we ran into the problem of inconsistent data types and divergent naming schemes. This can prove problematic and frustrating for team work at any one site, and particularly for the group as a whole.

In terms of lessons learned, the team members agreed that when one works closely with others, it is paramount to make sure that goals and interfaces are defined correctly the first time. To this end documentation certainly helps. Also, teams desiring to integrate their components with others should test them and the integration as far ahead of time as possible. Since we dealt with human-robot interactions and an interface, it is important to keep the naïve user in mind. Likewise, it was noted that at the event GRACE's verbal reactions were hard to understand. It is, therefore, important to have a speech generation component that is both easy to hear and understand.

Feedback to the human user is important for an easy to use interface. In the event, GRACE did not give feedback when aborting a movement and retrying the action; therefore, some of her actions were difficult to interpret. Likewise, as we mentioned earlier, a PDA interface failed to operate; however, no indication of the failure was given to the human user. Consequently, during GRACE's actual performance, the human user did not know why his PDA gestures were failing. In the event, GRACE did not give feedback when aborting a movement before its intended finish or retrying the action. However, GRACE constantly asked for feedback at the completion of actions: "Have I made it to the goal?" This was our way of determining that GRACE was finished completing an action or set of actions. However, to by-standers this repeated query made GRACE look dumber than she was. Designers of interfaces, therefore, must ensure that all interactions, including error-catching strategies appear "normal" and natural.

Acknowledgments

The research for this report was partly funded by the Office of Naval Research (N0001401WX40001) and the Defense Advanced Research Projects Agency, Information Processing Technology Office, Mobile Autonomous Robotic Software project (MIPR 02-L697-00).

Graphics and Photo Credits

The authors wish to thank the following for providing us with a photo record of the 2002 Robot Challenge: Figure 1, The City of Edmonton, Alberta, Canada; Figure 2, William Adams of the Naval Research Laboratory; Figures 3 and 4, Magda Bugajska of the Naval Research Laboratory; Figures 5-11, Adam Goode of Carnegie Mellon University.

General Thanks and a Disclaimer

The primary author of the paper would like to thank everyone on the team in working on AAAI 2002 Robot Challenge. He would also like to thank the NRL group for its cooperation and hard work in getting its portion of the task done for the event. In terms of this report, the primary author accepts full responsibility for any errors of omission or commission regarding the material contained herein.

Note

1. "GRACE: The Social Robot," http://www.palantir.swarthmore.edu/GRACE.

References

Anderson, J. R. and Lebiere, C. 1998. *The Atomic Components of Thought*. Lawrence Erlbaum: Mahwah, NJ.

Bruce, A., Nourbakhsh, I., and Simmons, R. 2002. "The Role of Expressiveness and Attention in Human-Robot Interaction," *Proceedings of the IEEE International Conference on Robotics and Automation*, Washington DC, May 2002.

Grosz, B. and Sidner, C. 1986. "Attention, Intentions, and the Structure of Discourse," *Computational Linguistics,* **12:3**, pp. 175-204.

Grosz, B., Hunsberger, L., and Kraus, S. 1999. "Planning and Acting Together," AI Magazine, **20:4**, pp. 23-34.

Huang, X.D., Alleva, F., Hon, H.W., Hwang, M.Y., Lee, K.F., and Rosenfeld R. 1993. "The SPHINX-II Speech-Recognition System: An Overview." *Computer Speech, and Language,* **7:2**, pp. 137-148.

Likhachev, M., and Arkin, R.C. 2001. "Spatio-Temporal Case-Based Reasoning for Behavioral Selection." In *Proceedings of the 2001 IEEE International Conference on Robotics and Automation*, (May 2001), pp. 1627-1634.

Nakauchi, Y. and Simmons, R. 2002. "A Social Robot that Stands in Line," *Autonomous Robo*ts. (May 2002), **12:3**, pp. 313-324.

Perzanowski, D., Schultz, A.C., Adams, W., Skubic, M., Abramson, M., Bugajska, M., Marsh, E., Trafton, J.G., and Brock, D. 2002. "Communicating with Teams of Cooperative Robots," *Multi-Robot Systems: From Swarms to Intelligent Automata,* A. C. Schultz and L.E. Parker (eds.), Kluwer: Dordrecht, The Netherlands, pp. 185-193.

Perzanowski, D., Schultz, A.C., Adams, W., Marsh, E., and Bugajska, M. 2001, "Building a Multimodal Human-Robot Interface." In *IEEE Intelligent Systems*, IEEE: Piscataway, NJ, pp. 16-21.

Perzanowski, D., Schultz, A.A., and Adams, W. 1998. "Integrating Natural Language and Gesture in a Robotics Domain." In *Proceedings of the International Symposium on Intelligent Control*, IEEE: Piscataway, NJ, pp. 247-252.

Wauchope, K. (1994) *Eucalyptus: Integrating Natural Language Input with a Graphical User Interface*, tech. Report NRL/FR/5510-94-9711, Naval Research Laboratory: Washington, DC.