From: AAAI Technical Report FS-92-02. Copyright © 1992, AAAI (www.aaai.org). All rights reserved.

# A Review of the First AAAI Robotics Competition

**R. Peter Bonasso**
Autonomous Systems Laboratory, the MITRE Corporation

**Thomas L. Dean**
Department of Computer Science, Brown University

**Abstract**

This paper summarizes the design, execution and results of the first AAAI Robotics Competition which took place in conjunction with the tenth national AI conference this past summer. As of this writing, no specific efforts have been made to evaluate the performance of the various robot intelligence techniques used in light of the scoring that took place. Rather this paper attempts to point out lessons learned and hints at the difficulties involved in attempting to stage and judge robot intelligence techniques working against a common set of tasks.

## Introduction

Robot applications, especially those in the service sector, require relatively sophisticated reasoning abilities, since service robots must cope with a far wider and less controllable range of situations than robots confined to factories. The recent advances in technology and anticipated changes in the demand for advanced robots led the American Association for Artificial Intelligence to sponsor a combined exhibition and competition to call attention to applications and technological challenges bearing directly on its members' expertise.

The exhibition/competition was designed to promote interaction among a wide range of industry and university researchers interested in these new applications of intelligent robotics. It featured robots, videos, and graphic images from university and industrial research laboratories around the world. The competition stressed the range of tasks robots must master to move about in an unstructured environment and cope with interference, both deliberate and inadvertent, from the humans with whom they share that environment. These robotics applications focus on apparently mundane tasks that are in fact every bit as complicated as more traditional AI tasks (e.g., playing master-level chess or providing decision support for air-traffic control) in terms of the interplay of behaviors and the physical interaction of the robot with the real world. These applications demand a degree of autonomy and robust execution unparalleled in prior commercial applications of AI technology.

## Background

The idea for the competition originated with Tom Dean inspired by a panel on household robots suggested by John Kender for the Ninth National Conference (AAAI-91).and ultimately chaired by Dean. Following the panel presentations, a startup company, Nomadics Technologies, demonstrated their robot wandering around in a crowd of fascinated attendees. From the response of the attendees and the press, it was clear that even the most cynical reporters and veteran researchers were not immune to the appeal of mobile robots.

At the time, AAAI was looking for ways to revitalize the national conference, to regain some of the excitement and enthusiasm of the early years. Dean used the surge of enthusiasm generated by the household robots panel to convince the AAAI that such a competition was feasible, that they should authorize a group to organize it, and, perhaps most importantly, that they should allocate AAAI funds to cover the costs. Dean envisioned a two phased event with the first phase being a safety/mobile competence demonstration, and the second phase involving two stages of environment exploration and mapping, and task execution based on the exploration. Pete Bonasso, Jim Firby, Dave Miller, and Reid Simmons had already begun discussions during AAAI-91 to refine these specifications.

It was certainly easy to imagine a competition that stressed more complicated forms of reasoning; the trick was to stage such a competition to occur within the temporal, environmental and political constraints imposed by a national conference. Conference attendees were used to seeing videos of robots performing in the lab, but often the performances were carefully orchestrated, used a variety of tricks to make the robot appear to be moving faster than it actually was, or simply documented one successful run out of many failed or less impressive runs. The competition would be held in public, in real time, and without the extensive controls possible in the laboratory. Most of those attending the competition would have high expectations.

## The Competition Tasks

The set of tasks had to be possible within the current state of the art, it had to exercise capabilities that were of

From: AAAI Technical Report FS-92-02. Copyright © 1992, AAAI (www.aaai.org). All rights reserved.

interest to the AI community, and it had to be both entertaining and fit in a three day venue. The first stage involved robots navigating in a cluttered environment and interacting with the people. This single stage comprised the first phase which served as a qualifying round to ensure that the robots were well enough behaved to participate in the subsequent events. In the second stage, the robots were to explore their environment and find and identify ten objects. In the third stage, the robots were to carry out commands issued by the judges, where a command consisted of a set of objects to visit in a specific order. The robots could make use of any information acquired in the second stage to expedite execution of these commands. The second and third stages together comprised the second and final phase of the competition. Prizes (1st through 5th place) were awarded for each of the two phases as well as for a variety of special categories suggested by the judges. The idea was to encourage machines that exercised common-sense capabilities involving planning, control, and rudimentary learning and spatial reasoning.

Careful selection of the tasks was also important because AAAI was concerned that the press not misrepresent the event. The organizing team endeavored to make it clear that our aims were to push the state of the art in autonomous agents and mobile robotics, educate the press and the AI community in the recent progress in the field, and provide an exciting educational opportunity for the participants that would result in a useful transfer of technology through a concerted effort on a common task.

A set of rules were drafted and iterated among the 30+ groups from national and international industry, government laboratories and academic institutions which showed early interest in participating. As the rules evolved, a scoring scheme resulted as follows:

- In stage one, points were allocated for autonomous navigation among stationary objects (a set of white cardboard boxes of various sizes), among moving objects (the judges), and points for endurance (a twenty-minute run).
- In stage two, points were allocated for autonomous exploration, autonomous perception of objects (8' PVC poles allowed to be rigged with sensor cues such as barcodes or IR beacons), and for maintaining the safe navigation exhibited in the first stage.
- In stage three, points were awarded again for autonomous navigation and perception, but also for using efficient schemes to accomplish the specified task and for speed of task completion.

The competition arena would consist of two 3' high hexagonal rings of foam-core walls roughly 70' x 70'. Since the large rings dwarfed the smaller robots, one of the rings was ultimately reduced in area by half.

## The Robot Entries

There were ten teams which finally fielded robots for the competition. These consisted of four small robots -- TJ from IBM, Chip from the University of Chicago, Odysseus from CMU, and Huey from Brown University; and six large robots -- Uncle Bob from MITRE, Carmel from University of Michigan, Soda-Pup from NASA-JSC, SRI's Flakey, Scarecrow by Dave Miller of JPL on sabbatical to MIT, and Buzz from the Georgia Institute of Technology. A team from the San Francisco Robotics Society had plan to enter, but ended up as a demonstration only.

There was remarkable agreement on the basic form of the robot intelligence architectures. Almost every architecture consisted of a multi-level control system coupling low-level routines for sensing, primitive navigation (e.g. wall following) and obstacle avoidance operating continuously in highly responsive feedback loops, and high-level routines for planning, and map generation, relying on the low-level routines to keep the robot in synch with real-time events. There were systems based on variants of Brooks' subsumption architecture (IBM ) and on Rosenschein and Kaelbling's situated automata approach (MITRE). There were also systems that did not adopt any particular architecture but incorporated the lessons that such architectures have taught us over the years. For instance, the Brown entry was a modular object-oriented software system, employing multiple processors to manage low-level and high-level routines asynchonously. In general, the software developed for this competition was extraordinary for its sophistication and ease of modification. In some cases, code was considerably revised during the competition to cope with hardware failures and complications introduced by conditions in the conference hall.

## The Competition

For the participants and organizers, the competition began Sunday July 12 when the shipping crates were unpacked and the arena constructed in the San Jose Convention Center. Inevitably, there was equipment damaged in transit, tools and parts that had been left behind, software complications for those using borrowed computing machinery, and a wide variety of complications having to do with the arena and convention hall. With regard to this last, the harsh, uneven lighting posed problems to vision systems, and, perhaps the most important factor for many of the competitors, the hall was a veritable soup of rf noise: portable microphones, transmitters used by the press, two-way radios used by the convention center employees, dimmers and starters for the powerful halogen lighting, not to mention that the competitors used a wide range of incompatible devices for transferring data between robots and remote computing equipment. Problems with communication and equipment damaged in transit would figure prominently in the outcome of the competition.

During the two days prior to the competition, hardware was repaired and replacement equipment flown in. Software underwent major revisions throughout the competition. There was an enormous amount of sharing of expertise and equipment; the competitors helped one another in a myriad of ways. The fact that they understood the tasks in intimate detail made communication almost

From: AAAI Technical Report FS-92-02. Copyright © 1992, AAAI (www.aaai.org). All rights reserved.
telepathic; all of the competitors wanted to know exactly how the others had approached the competition. Some used very general approaches that could deal with a wide variety of tasks; others adopted a very task-specific approach. Some had very primitive hardware but very sophisticated software and some relied on hardware specifically designed for the task at hand. Indeed the experience was exhilarating for all the competitors because for 5 days there was assembled under one roof some of the brightest, most intelligent and incredibly clever crowd of hardware and software hackers from across the nation.

## Roaming in a Semi-structured Environment

The first stage, conducted on the first day, was meant to exercise the robots' abilities to navigate autonomously avoiding obstacles both fixed and moving. The judges were encouraged to interact with the robots. To the teams' consternation the judges decided to push their capabilities to the limit, both to determine some criteria for distinguishing among the competitors, and to better understand the limits of the technical approach used. Ken Forbus, Ben Kuipers, Larry Matthies, Stan Rosenschein, Yoav Shoham, and Beverly Woolf were the judges for the first day, dashing in front of the robots, moving boxes, and even in some cases lying down in front of the robots. A number of the robots were cheered for their graceful movements, their agility, and their various means of communicating with the judges and their handlers. Computer generated sounds and voice synthesizers were particularly appreciated by the crowd. While there were some problems involving robots communicating with remote computing devices via radio modems on the first day, many of the robots were able to operate autonomously for this stage, using remote computing not at all or only for initial start up or periodic updates for graphic displays or to some sort of high-level planner.

The attendees were captivated by one of the competitors, Jacob Milstein, the son of Dave Miller who together built an intriguing special-purpose machine designed just for this competition -- Scarecrow. Scarecrow, was a real crowd pleaser but did not fare particularly well in the first stage as its primary means of sensing involved crashing into to things and people (deemed "obstacle avoidance by collision detection" by one observer) and then dashing off in the opposite direction.

The winner of the first stage was IBM 's TJ which combined ultrasonic sensors for long-range obstacle detection with near-infrared sensors for short-range obstacle detection for a robust, agile, navigation and obstacle avoidance system. SRI's Flakey was a close second followed by Michigan's Carmel, Georgia Tech's Buzz, and Chicago's Chip, filling out the top five ranking robots. The other competitors finished in order from sixth through last, Miller and son's Scarecrow, NASA's Soda Pup, Brown's Huey, CMU's Odysseus, and MITRE's Uncle Bob. Because all of the robots proved to be safe and to smoothly negotiate the obstacles after about ten minutes, the single largest discriminant among the judges became the behavior of the robots as they were being contained and otherwise harassed by the judges.

In the hours between the end of the first day's activities and the beginning of the second day's, there were many minor and major hardware disasters. Several teams made frantic calls to manufacturers and colleagues back home to ship replacement parts via overnight shipping companies. There were still more discoveries of damage in transit. One particularly poignant story involved MITRE's Uncle Bob which was discovered to have a wheel assembly badly bent out of alignment. The team was able to compensate by modifying their software to some extent; however, in their exhaustion later that night, one of the team members accidentally shorted out a part that resulted in freezing the top rotating sensor platform. Uncle Bob missed the second day but managed to return on the last day to finish the competition despite the damaged drive assembly and the incapacitated head. This feat pointed out the importance of software in being able to overcome some severe hardware deficiencies, albeit at the cost of greatly reduced performance.

A critical vision board on Chicago's Chip had burned out and the team was re-coding furiously in an effort to get get around the problem. They were also awaiting the delivery of a new board in the hope that a replacement board would correct their problems. The next day would be even more frustrating for the Chicago team when the replacement board arrived and promptly burned out just like the earlier one. After an extraordinary effort Chicago had to withdraw from the competition entirely.

## Exploration in a Semi-Structured Environment

On the second day, stage two was held, wherein the robots were to explore the rings and find and locate the objects represented by the PVC poles. Robots that were not able to identify labels on objects were forced to first differentiate a pole from a box and then distinguish each object from the others using object location. With the exception of Flakey, robots that attempted this strategy turned out to be far too slow to compete with the other robots which used specific stimuli attached to the poles to aid in identification. Also, robots with cameras on stalks or tall robots that could see over the boxes excelled while those relying on local sensing were at a disadvantage. Michigan, CMU, Georgia Tech and MITRE all employed cameras that allowed them to perceive objects from some distance over the boxes.

This event was meant to encourage teams to construct some sort of an internal representation of the ring, encoding the location of the objects for use in the final stage of the contest in which the judges would ask the robots to visit a set of objects in some particular order. Most of the robots did indeed build some sort of map of the ring and the judges were shown graphical displays depicting these maps during the competition.

The judges on the second day consisted of Jerry Dejong, Mike Georgeff, Jim Hendler, Ian Horswill, Matt Mason, and Martha Pollack. The judges worked hard to deal with

From: AAAI Technical Report FS-92-02. Copyright © 1992, AAAI (www.aaai.org). All rights reserved.
disparities in the size, speed, and sensor capabilities of the various robots. A large part of the judges' discussions centered around what constituted an object classification. Some robots would clearly identify the objects at a distance and then navigate to them to establish the object location, while Scarecrow, for example, couldn't identify an object until it ran into it. After the tallies of the second stage, Carmel (U Mich) was on top followed by Buzz (Ga Tech), Flakey (SRI), Odysseus (CMU) and Scarecrow (Miller) for the top five. Soda-pup (JSC) was sixth followed by TJ (IBM) and Huey (Brown).

TJ was hampered by a broken analog-to-digital converter that crippled their pole recognition system. Nevertheless, the IBM team members were able to modify their software so that TJ found and identified three objects.

Brown was one of the entries to rely entirely on sonar for sensing the environment. Huey, a small RWI-based robot, only had eight sonars, so despite a sophisticated technique for interpreting sonar data and positioning to obtain multiple views for pole recognition, their robot was not able to recognize any of the poles during the allotted time of twenty minutes. Flakey, on the other hand, successfully demonstrated that sonar could be used to pick up pole indicators as a backup to the primary sensor (structured light).

The surprise of the second day was Scarecrow which managed to find and detect seven of the ten objects in the allotted time. As could be expected by its pseudo-random walk, the robot found the first four objects in five minutes and spent the remainder of the 20 minute period looking for the rest. Scarecrow lurched about the ring always seeming on the brink of crashing to the floor, smashing into boxes and occasionally brushing up against a pole with its tall circular antennae that was used to recognize poles. Miller and son had rigged each pole with a steel wool ring at exactly the height of the circular antennae. If scarecrow brushed up against a pole in the right way, the steel wool would serve to close an electrical connection involving the circular antennae allowing Scarecrow to register that it had located a pole. Scarecrow offered a dramatic illustration of a theoretical result that a short random walk in an undirected graph will ultimately visit every location in the graph with high probability.

Operating only on relays, Scarecrow could not remember anything during its exploration of the ring. This and the fact that several other robots found all ten objects would put it at somewhat of a disadvantage in the final stage of the competition in which it would be asked to visit specific locations. Scarecrow would have to use the same random walk strategy to carry out the judges' request. However, what scarecrow lacked in intelligence (like its namesake in the Wizard of Oz, Scarecrow had no brains) it made up for in raw speed and as a result scarecrow was still a contender going into the final event.

## Carrying Out a Directed Task in a Semi-Structured Environment

The judges on the final day consisted of James Crawford, Erann Gat, Steve Hanks & Henry Kautz (sharing duties because of other conference conflicts), Maja Mataric, Nils Nilsson, Howie Shrobe. Again they had the extraordinarily difficult job of comparing the performance of very different robots, performing in different rings, and in some cases carrying out subtly different tasks. A particular problem concerned specifying the task for each robot in stage three. The general task was for the robot to navigate to and re-locate three objects in an order given by the judges and then to return home. The issue was how to fairly compete robots which only found a few objects in stage two with those which had found all ten. Since most of the robots which only found a few poles ran in the small ring, the answer was to use a common subset of poles for the second ring robots and vary the order, and to adjust the task poles for the robots in the larger ring based on the location of the poles that had been found.

Eight teams competed on the third day, including MITRE which had built a stage two map earlier that morning. As discussed previously, Huey and Chip did not compete. The final event would be run in two heats with the robots performing in reverse order of their standings on the previous day. TJ recovered from its low ranking on the previous day with a fast run, dead-reckoning on the map that it built on the previous day. Scarecrow ran a close second to TJ, randomly visiting many poles, but using a mechanical state counter to indicate when it found a desired pole in the desired order. Soda-Pup and Uncle Bob put in credible performances but were not in contention for the top places. Two of the top four contenders from the previous day, Odysseus and Flakey, turned in impressive performances with both of them quickly completing the specified tasks if a little slower than TJ and Scarecrow.

Georgia Tech faltered after its second ranked performance on the previous day. There were a number factors likely involved, but rf interference was the primary suspect. As the excitement regarding the competition grew over the three days the press had gathered with their array of electronic devices. It was nearly impossible to control all those trying to record the event, and Buzz relied on communication with a remote computer for success.

The unabashed intrusion of the press into the rings during every stage of the competition was completely unforeseen. They were obviously unaware that they were recognized as obstacles to be maneuvered around and may have changed the performance of some of the slower robots, though the top five finalists all coped well with moving obstacles in stages two and three.

The last robot to run, Michigan's Carmel, and the leading contender from the previous day started with some problems. There was a false start in which one of the team members entered the wrong command and the robot confused everyone by flawlessly executing a sequence quite different from that given by the judges. On its second

From: AAAI Technical Report FS-92-02. Copyright © 1992, AAAI (www.aaai.org). All rights reserved.
attempt, Carmel seemed to get disoriented and at one point stood by the wall of the ring scanning the crowd for the obstacle it was looking for. It was as though an Olympic champion had advanced to the finals leading in the standings only to fail inexplicably in the last seconds. In this case, however, Carmel recovered. The judges gave the robot one final try in which it executed the sequence flawlessly turning in the fastest time. This performance coupled with that of the previous day led the judges to award Carmel first place in the second phase.

Giving robots additional tries within the twenty minute period in any of the stages was not part of the original rules but was unanimously adopted by all judges on all three days. As Nils Nilsson put it "we've all given demos before".

### And the Winner . . .

The final standings for the second phase were: Carmel with 544 points, Flakey with 452 points, Odysseus with 449 points, Scarecrow with 413 points, Buzz with 397 points, Soda -Pup with 373 points, and TJ with 372 points.

In the closing ceremonies, Pat Hayes presented the awards. In addition to the top five places in the two phases, MITRE was awarded a certificate for its dramatic recovery, Johnson Space Center for its impressive performance with the shortest development time, and Miller and son for their innovative design.

## Summary of Findings

At a fundamental level two preliminary claims could be made based on the results of the competition. One, it appears that a variety of techniques are known which allow a mobile robot to be safe while navigating for whatever task among objects with continuous z-extents and among people. Two, that robot intelligence architectures have advanced to the point where several technical approaches can be successfully applied in a non-laboratory environment to a common task not specified by advocates of the approaches.

There are more specific results which while perhaps not surprising, nonetheless were borne out by the competition. Robots that used simple sensors and sophisticated methods for dealing with uncertainty in sensing and movement performed the tasks at an unacceptably slow rate. As was apparent from Scarecrow's performance, sophisticated reasoning was not necessary to perform passably for some types of tasks. Carmel's accurate dead reckoning and long-range sensing were significant factors in the Michigan team's success, and suggest that quite robust robot behavior can be obtained if the task lends itself to those facets of the robot hardware suite.

The rf problems need to be dealt with soon and with concerted efforts if we are ever to realize multiple robots working in the same environment. Here is an area where collaboration with the communications/signal processing disciplines should yield much fruit.

The popular appeal of mobile robots working alongside humankind is strong and continues to increase. It seems embedded in our collective psyche for reasons that can only be surmised. But that appeal ought to encourage researchers to participate in more events such as this one to continue to educate the general community as to how far we have really come, and to counter the mistaken notions of science-fiction motion pictures.

Finally, we believe that the tension between approaches tailored to a particular problem and those striving for generality is extremely healthy for the field. The results of this tension will help us understand what robotic AI technology can be transitioned to the real world now and for what kind of tasks, versus that which still requires more thought and experimentation. Future competitions should continue to play these very different approaches against one another in an effort to better understand the basic tradeoffs.