AI Magazine Volume 18 Number 1 (1997) (© AAAI)

Many Robots Make Short Work

Report of the SRI International Mobile Robot Team

Didier Guzzoni, Adam Cheyer, Luc Julia, and Kurt Konolige

■ Indoor mobile robots are becoming reliable enough in navigation tasks to consider working with teams of robots. Using SRI International's open-agent architecture (OAA) and SAPHIRA robot-control system, we configured three physical robots and a set of software agents on the internet to plan and act in coordination. Users communicate with the robots using a variety of multimodal input: pen, voice, and keyboard. The robust capabilities of the OAA and SAPHIRA enabled us to design and implement a winning team in the six weeks before the Fifth Annual AAAI Mobile Robot Competition and Exhibition.

At the SRI International Artificial Intelligence Lab, we have a long history of building autonomous robots, from the original SHAKEY (remember the STRIPS planner?) through FLAKEY (Congdon et al. 1993) and, more recently, the Pioneer class of small robots. Our current research focuses on real-time vision for robots and multirobot planning using an agent-oriented architecture.

For the Fifth Annual AAAI Mobile Robot Competition and Exhibition, held as part of the Thirteenth National Conference on Artificial Intelligence (AAAI-96), we wanted to showcase our research, especially the ability to control multiple robots using a distributed set of software agents on the internet. The agent technology, called the *open-agent architecture* (OAA), was developed at SRI as a way of accessing many different types of information available in computers at different locations.

Office Navigation

In the Office Navigation event, a robot starts from the director's office, determines which of two conference rooms is empty, notifies two professors where and when the meeting will be held, and then returns to tell the director. Points are awarded for accomplishing the different parts of the task, communicating effectively about its goals, and finishing the task quickly. Our strategy was simple: Use as many robots as we could to cut down on the time to find the rooms and notify the professors. We decided that three robots was an optimal choice: enough to search for the rooms efficiently but not too many to get in each other's way or strain our resources. We would have two robots searching for the rooms and professors and one remaining behind in the director's office and tell him/her when the meeting would be. We were concerned that leaving one robot behind as a mobile telephone was stretching things a bit; so, we cleared our strategy with the judges well before the competition.

The two search robots are Pioneer-class robots, portable robots first developed by SRI for classroom use and now manufactured by Real World Interfaces, Inc. (RWI). They run SRI's SAPHIRA control software, which navigates the robots around an office environment, keeping track of where they are using perceptual cues from the robot sensors. Each Pioneer robot has seven sonar sensors, a fast-track vision system from Newton Labs, and a portable computer on top with a radio ethernet for communication to a base station (figure 1). The fast-track system is an interesting device: It consists of a small color video camera and a low-power processor. It can be trained to recognize particular colors and will indicate the position of any object with this color. We decided to use the vision system to

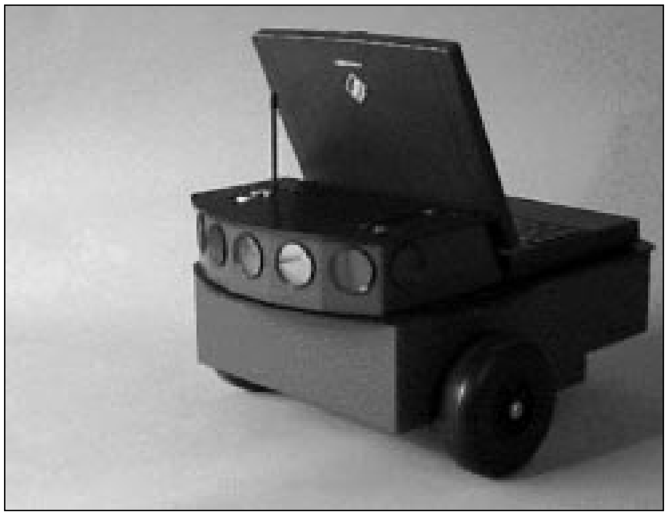


Figure 1. A Pioneer Robot with Laptop Host Computer.

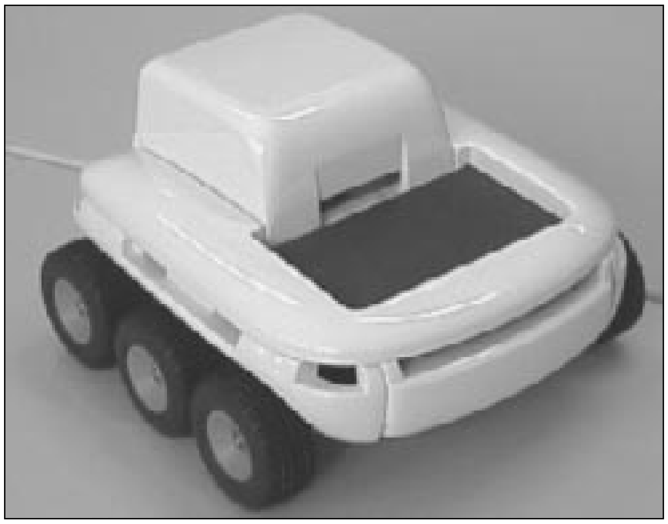


Figure 2. A Koala Robot from the Swiss Federal Institute of Technology.

find people in the conference rooms and trained it to recognize red. If the judges would wear red shorts, the vision system would easily pick them out.

The robot in the director's office didn't have to move, just relay information to the director. We used a Koala robot, a small, six-wheeled vehicle under development at the Swiss Federal Institute of Technology (figure 2). The Koala has infrared sensors, which enable it to avoid obstacles but make it difficult to map the environment because they do not give a reliable range estimate. We just kept the Koala stationary, with a portable PC on top to communicate with the other robots and talk to the director.

Each robot, by virtue of the radio ethernet, is a node on the internet and an agent in the OAA. Other agents, residing on the base station, included a database agent for holding information relayed back by the robots, a mapping agent for determining and displaying where the robots are, a strategy agent for coordination, and interface agents (speech, pen gestures) for giving the robots commands. If we had a connection to the outside world, we could have run the robots from anywhere in the world!

One of the interesting aspects of the agent architecture is that the robots are capable of a good deal of autonomy. For example, the strategy agent might tell them where to go and even give them a path. The robots are responsible for navigation, avoiding obstacles and finding the correct goal position. If they fail, they contact the strategy agent about the problem and wait for instructions. The connection between the robots and the rest of the agents can be low bandwidth.

This was our strategy; now we had to execute it. We arrived at the contest late, during the preliminary rounds, and hastily set up our base station and put in the map of the office environment. All did not go smoothly—several unanticipated coordination problems between the mapping agent and the robots caused us some frustrating moments. We hadn't realized the planner would happily plan paths through rooms with two doors, something the robots didn't like because they get lost easily in rooms and do much better in the corridors. One of our robots was injured in a fall as we were packing up at SRI; fortunately, RWI was there and loaned us another Pioneer. Late in the evening of the second day, we had a perfect preliminary-round run. Figure 3 is a diagram of the paths of the two Pioneer robots. Each of them started out going to a different conference room. The solid-line robot arrived at the first room, found it occupied, and then started heading for a professor's office. Meanwhile, the dotted-line robot reached the second conference room and, after a short time, decided there were no people present. It then went to the second professor's office.

The solid-line robot arrived at the professor's office just a little ahead of the dotted-line robot. As the robots entered the professors' offices, they announced the time of the meeting and the conference room. At this point, the Koala robot announced the meeting to the director, and the task was completed.

The next day, the finals, was almost an anticlimax, except for the large audience and the presence of Alan Alda and the camera crew from *Scientific American Frontiers.* We started the Pioneers out from the director's office, and they went happily on their way. There was too much going on to keep track of both of them: They were both announcing what they were doing, the camera folks were dancing around, and we had to tell the audience what was happening. Before we knew it, both robots were in the professors' offices, announcing the meeting: only 4 minutes and 30 seconds to completion! After watching the other teams, we knew the nearest time would be close to 10 minutes. Parallelism does work sometimes....

In the rest of this article, we briefly describe SAPHIRA (the robot control program) and the OAA, and then discuss our implementation of multirobot planning and control using these tools.

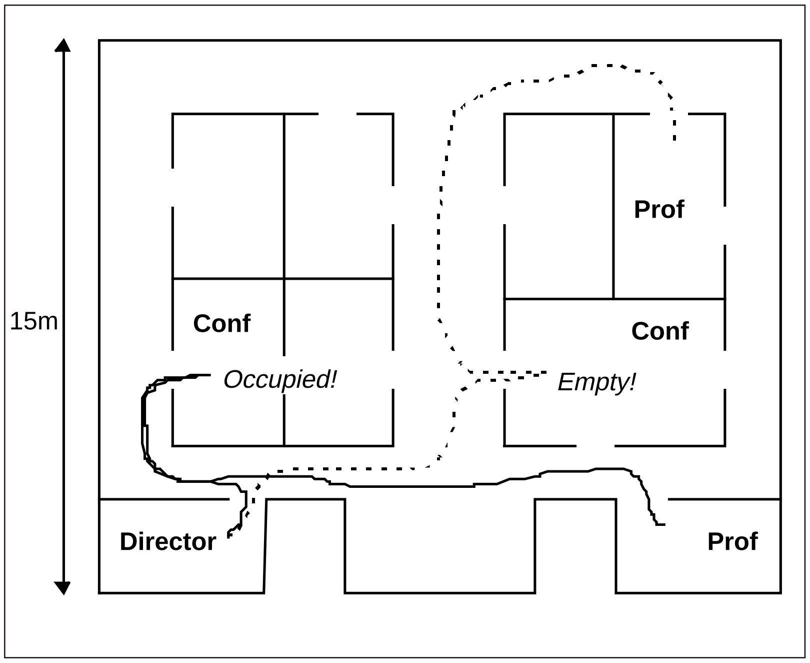


Figure 3. Paths of the Two Pioneer Robots.

SAPHIRA and the Open-Agent Architecture

In this section, we discuss the SAPHIRA robot controller and the OAA.

The SAPHIRA Robot Controller

The SAPHIRA architecture (Konolige 1997; Saffiotti, Konolige, and Ruspini 1995) is an integrated sensing and control system for robotics applications. At the center is the *local perceptual space* (LPS) (figure 4), a geometric representation of space around the robot. Because different tasks demand different representations, the LPS is designed to accommodate various levels of interpretation of sensor information as well as a priori information from sources such as maps. Currently, the major representational technologies are (1) a grid-based representation similar to Moravec and Elfes's occupancy grids (Moravec and Elfes 1985) built from the fusion of sensor readings; (2) more analytic representations of surface features such as linear surfaces, which interpret sensor data relative to models of the environment; and (3) semantic descriptions of the world, using structures such as corridors or doorways (*artifacts*), which are the product of a bottom-up interpretation of sensor readings or a top-down refinement of map information.

The LPS gives the robot an awareness of its immediate environment and is critical in the tasks of fusing sensor information, planning local movement, and integrating map information. The perceptual and control architecture make constant reference to the LPS. One can think of the internal artifacts as SAPHIRA's beliefs about the world, and most actions are planned and executed with respect to these beliefs.

In Brooks's (1986) terms, the organization is partly vertical and partly horizontal. The vertical organization occurs in both perception (left side) and action (right side). Various perceptual routines are responsible for both adding sensor information to the LPS and processing it to produce surface information that can be used by object recognition and navigation routines. On the action side, the lowest-level behaviors look mostly at occupancy information to do obstacle avoidance. The basic building blocks of behaviors are *fuzzy rules,* which give the robot the ability to react gracefully to the environment by grading the strength of the reaction (for example, turn left) according to the strength of the stimulus (for example, distance of an obstacle on the right). Navigation routines make use of map information to guide the robot toward goal locations, for example, to a corridor junction. At the same time, registration routines keep track of sensed objects, constantly relating them to internal map objects to keep the robot accurately positioned with respect to the map. Thus, SAPHIRA is able to accept a *plan,* a sequence of way points to a final goal, and execute it while the robot is kept localized within the global map.

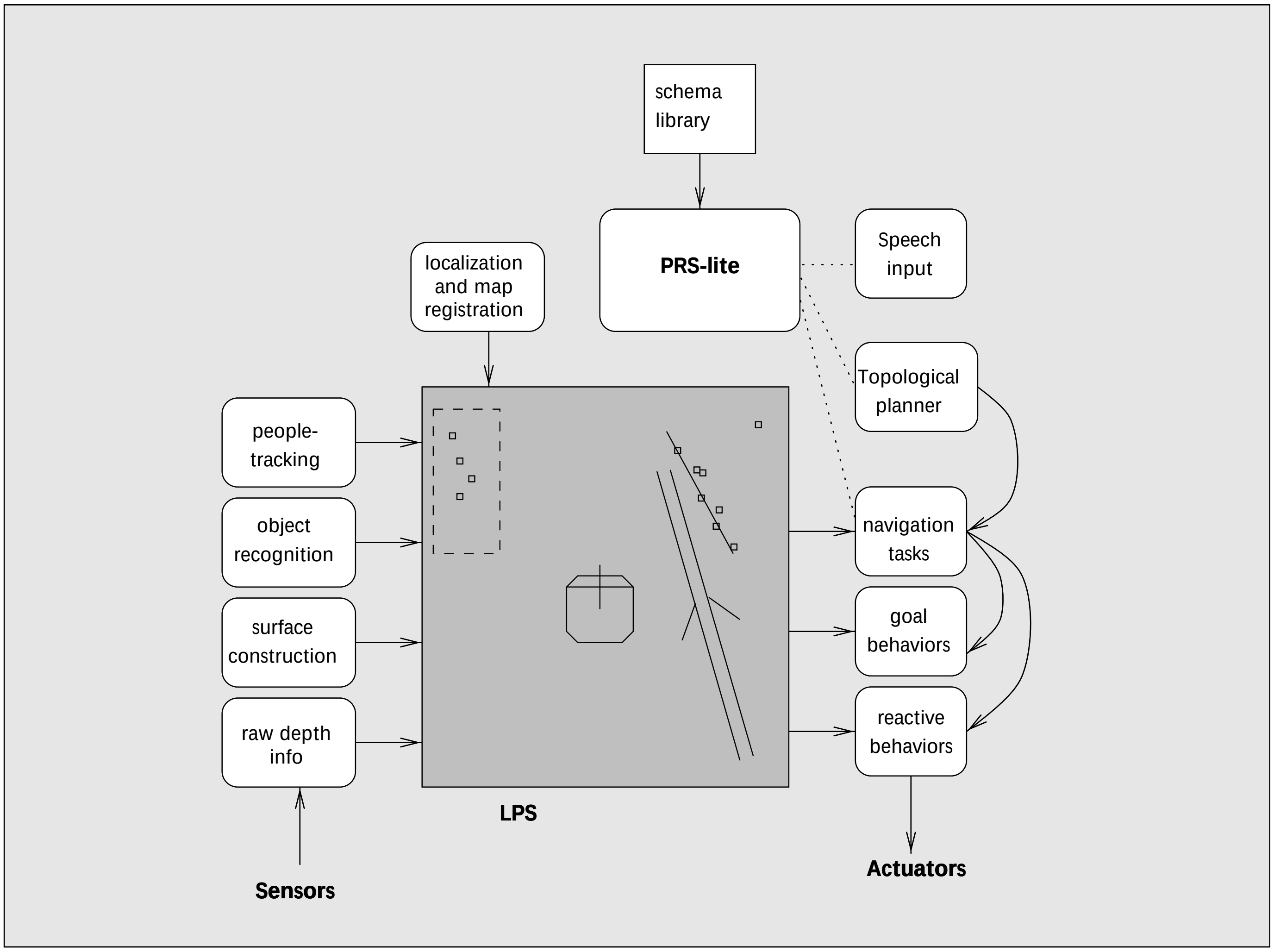


Figure 4. SAPHIRA *System Architecture.*

Perceptual routines are on the left, action routines on the right. The vertical dimension gives an indication of the cognitive level of processing, with high-level behaviors and perceptual routines at the top. Control is coordinated by the procedural reasoning system (PRS-LITE), which instantiates routines for task sequencing and monitoring and perceptual coordination.

The Open-Agent Architecture

When planning our strategy for how to approach this year's robot contest, we decided to take advantage of our recent integration of SAPHIRA as an agent within the OAA (Cohen et al. 1994). The OAA is a framework for constructing multiagent systems that has been used by SRI and clients to construct more than 15 applications in various domains (Moran et al. 1997; Kameyama, Kawai, and Arima 1995). Applying the OAA to the Office Navigation event in the robot competition could provide the following advantages:

Distributed: Agents can run on different platforms and operating systems and can cooperate in parallel to achieve a task. Some agents could be placed locally on each robot's laptop controller, but other services could be stored on a more powerful workstation.

Plug and play: Agent communities can be formed by dynamically adding new agents at run time. It is as easy to have multiple robots executing tasks as it is to have just one.

Agent services: Many services and technologies encapsulated by preexisting agents can easily be added as resources provided by our agent community. Useful agents for the robot domain would include database agents, mapping agents, and agents for text to speech, speech recognition, and natural language, all of which are directly reusable from other agent-based applications.

Multimodal: The agent architecture has been designed with the human user in mind (Cheyer and Julia 1995). Agents have been

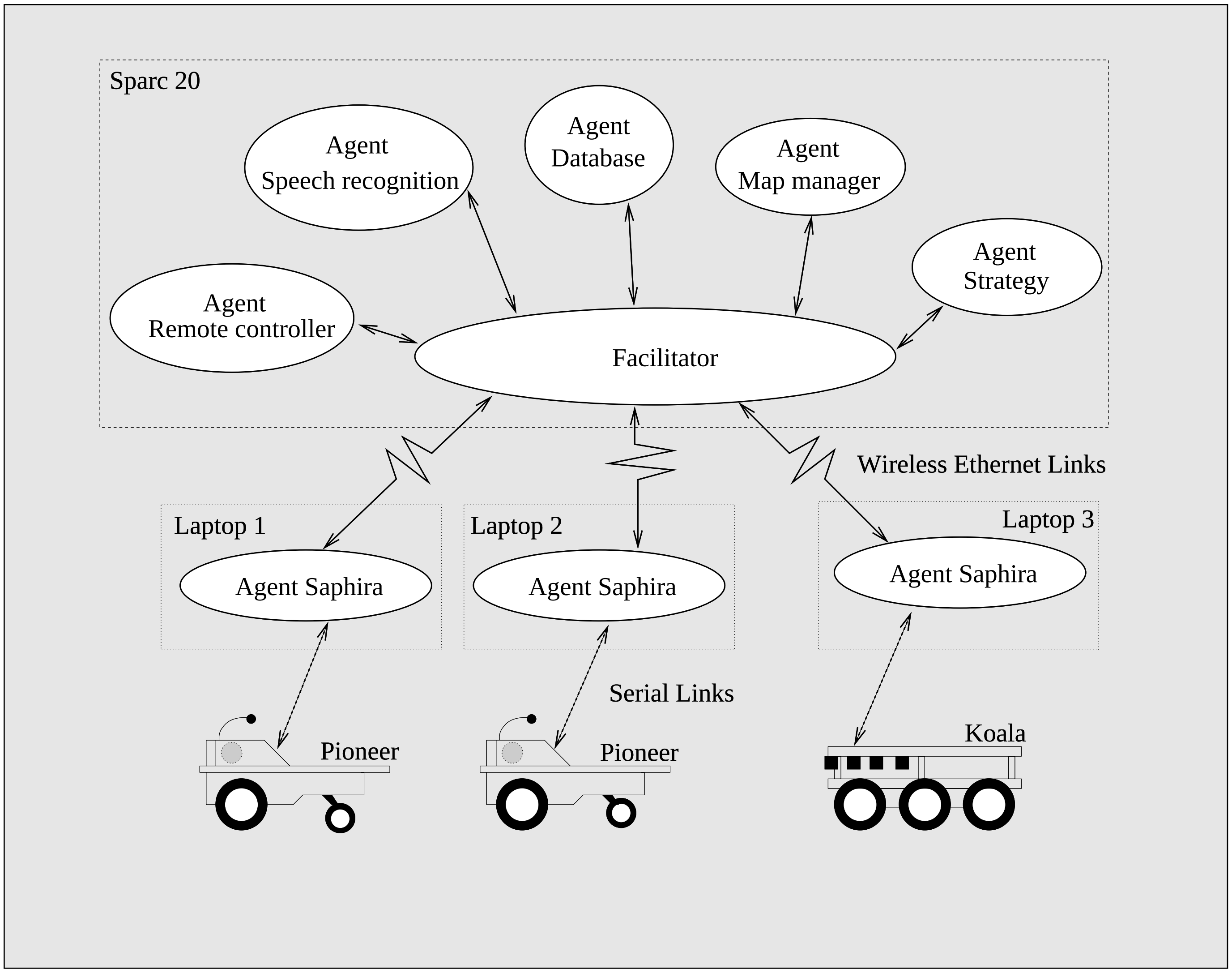


Figure 5. Organization of Physical and Software Agents for the Robot Contest.

developed to allow people to naturally combine, drawing, speaking, and writing with more standard graphic user interface approaches when addressing the set of distributed agents. In a robot domain, we can monitor the progress of robots on a map and, if required, give them instructions by speaking—"you are here, facing this direction" (while drawing an arrow) or "pick up this object" (while indicating a target using circles, pointing, or arrow gestures).

Mobile: The agent libraries are lightweight enough to allow multiple agents to run on small, wireless personal desktop assistants or laptops, and communication among agents is fast enough to provide real-time response for the robot domain.

The OAA uses a distributed architecture in which a *facilitator agent* is responsible for scheduling and maintaining the flow of communication among a number of client agents. Agents interact with each other through an *interagent communication language* (ICL), a logic-based declarative language based on an extension of Prolog. The primary job of the facilitator is to decompose ICL expressions and route them to agents that have indicated a capability of resolving them. Because communication occurs in an undirected fashion, with agents specifying what information they need, not how this information is to be obtained, agents can be replaced or added in a plug-and-play fashion.

Each agent in the OAA consists of a wrapper encapsulating a knowledge layer written in Prolog, C, Lisp, Java, Visual Basic, or Borland's Delphi. The knowledge layer, in turn, might lie on top of existing stand-alone applications and serves to map the functions of the underlying application into the ICL. In the case of the physical robots, we installed an agent interface on top of SAPHIRA, so that information about the robot's location and commands to navigate the robot were made available to all agents.

The OAA agent library provides common functions across all agents. Each agent can respond to or produce requests for information or service and can install triggers to monitor real-world conditions. Triggers can refer to temporal events, changes in local or remote data values, specific agent communication messages, or domain-specific test conditions provided by some agent (for example, a trigger request "when mail arrives from Bob..." will automatically be installed by the facilitator on the mail agent, which can perform this verification).

The system we developed is made of a set of independent agents (including robots) that are able to communicate to perform cooperative tasks.

Robots as Physical Agents

The system we developed is made of a set of independent agents (including robots) that are able to communicate to perform cooperative tasks. An operator can graphically monitor the whole scene and interactively control the robots. A top-level program, the *strategy agent,* was designed to synchronize and control the robots and software agents.

Figure 5 is a diagram of the complete system, including the physical location of all agents. The facilitator, database agent, map-manager agent, strategy agent, and speech-recognition agent were running on a UNIX workstation (SPARC20). On the robots, each SAPHIRA agent was running (under WINDOWS 95) on a laptop computer, each equipped with sound devices and text-to-speech converters. The link between the robots and the SPARC20 was through wireless ethernet links.

All agents start running and connect to the facilitator, registering their capabilities so that other agents can send them requests. This part of the agent architecture is essential: Agents must be able to access each other's capabilities in a uniform manner. Many of the interface agents already exist at SRI, the speech-recognition and pen-gesture–recognition agents, for example. To access these capabilities for the robots, we have only to describe how the output of the interface functions should invoke robot commands. In addition, because agents are able to communicate information by asking and responding to queries, it is easy to set up software agents, such as the mapping and strategy agents, to track and control multiple robots. We briefly describe the capabilities of the agents.

Robot Information and the Database Agent

Each *robot agent* provides information about the robot state and accepts commands to control the robot. The information includes position with respect to the robot's internal coordinate system; robot movement status—stopped, moving forward, turning; and currently executing behaviors on the robot. An interesting problem is how two agents maintain a consistent coordinate system. Commands that are robot relative, for example, move forward, are interpreted with respect to the robot's internal coordinate system. Other commands, such as "go to office EK288," must be interpreted with respect to a common global framework. The *database agent* is responsible for maintaining a global map and distributing this information to other agents when appropriate. Each physical robot has its own copy of the global map, but these copies need not be exactly alike. For example, an individual map can be missing information about an area the robot has no need to visit.

During movement, each robot keeps track of its global position through a combination of dead reckoning (how far its wheels have moved) and registration with respect to objects that it senses. It communicates with the database agent to update its position about once a second and report any new objects that it finds so that they can be incorporated into the global database and made available to other agents. In this way, the database agent has available information about all the robot agents that are currently operating.

The Mapper Agent and Multimodal Input

If a robot becomes lost, it can query the facilitator to help relocalize. Currently, this querying means human intervention: The facilitator signals that a particular robot is lost and asks for a new position for the robot. The state of each robot is displayed by the *map-manager agent,* or *mapper.* All currently known objects in the database, as well as the position of all robots, are constantly updated in a two-dimensional window managed by this agent. Figure 6 shows the mapper's view of the database contents. Corridors, doors, junctions, and rooms are objects known to the mapper. A robot's position is marked as a circle with an arrow in it, showing the robot's orientation.

To correct the position of a lost robot, the

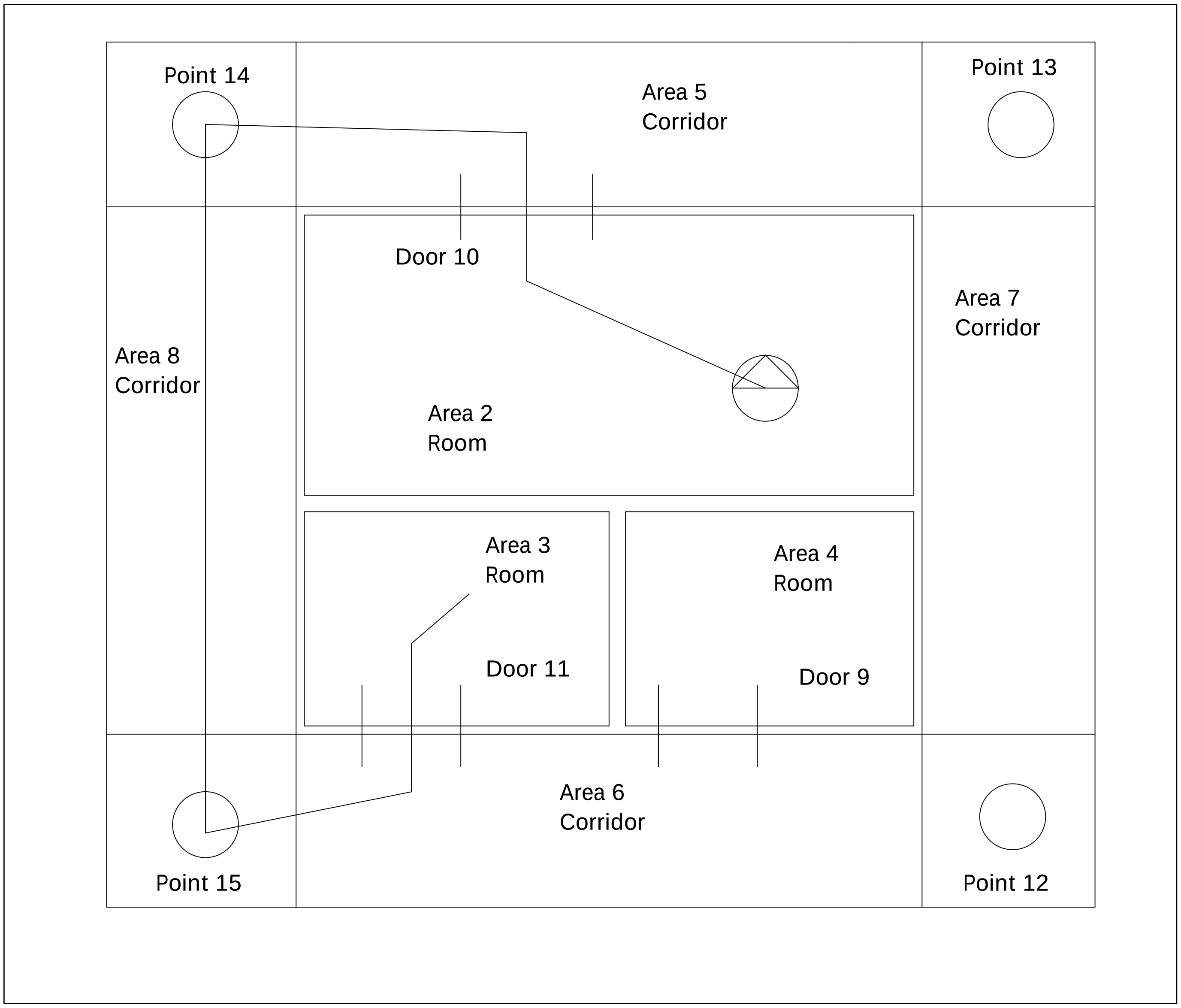


Figure 6. The Mapping Agent's View of the Database.

user can point to a position on the map where the robot is currently located or simply can describe the robot's position using speech input. This feature is one of the most useful of the OAA architecture—the integration of multimodal capabilities.

Currently, the system accepts either voice input or pen gestures. The interpretation of the gestures depends on the context. For example, when the robot is lost, the user can tell it where it is by drawing a cross (for the location) and an arrow (to tell the robot where it faces) on the map. Using two-dimensional gestures in the human-computer interaction holds promise for recreating the paper-pen situation where the user is able to quickly express visual ideas while he/she is using another modality such as speech. However, to successfully attain a high level of human-computer cooperation, the interpretation of online data must be accurate and fast enough to give rapid and correct feedback to the user. The gesture recognition engine used in our application is fully described in Julia and Faure (1995). There is no constraint on the number of strokes. The latest evaluations gave better than 96-percent accuracy, and the recognition was performed in less than half a second on a PC 486/50, satisfying what we judge is required in terms of quality and speed (Moran et al. 1997).

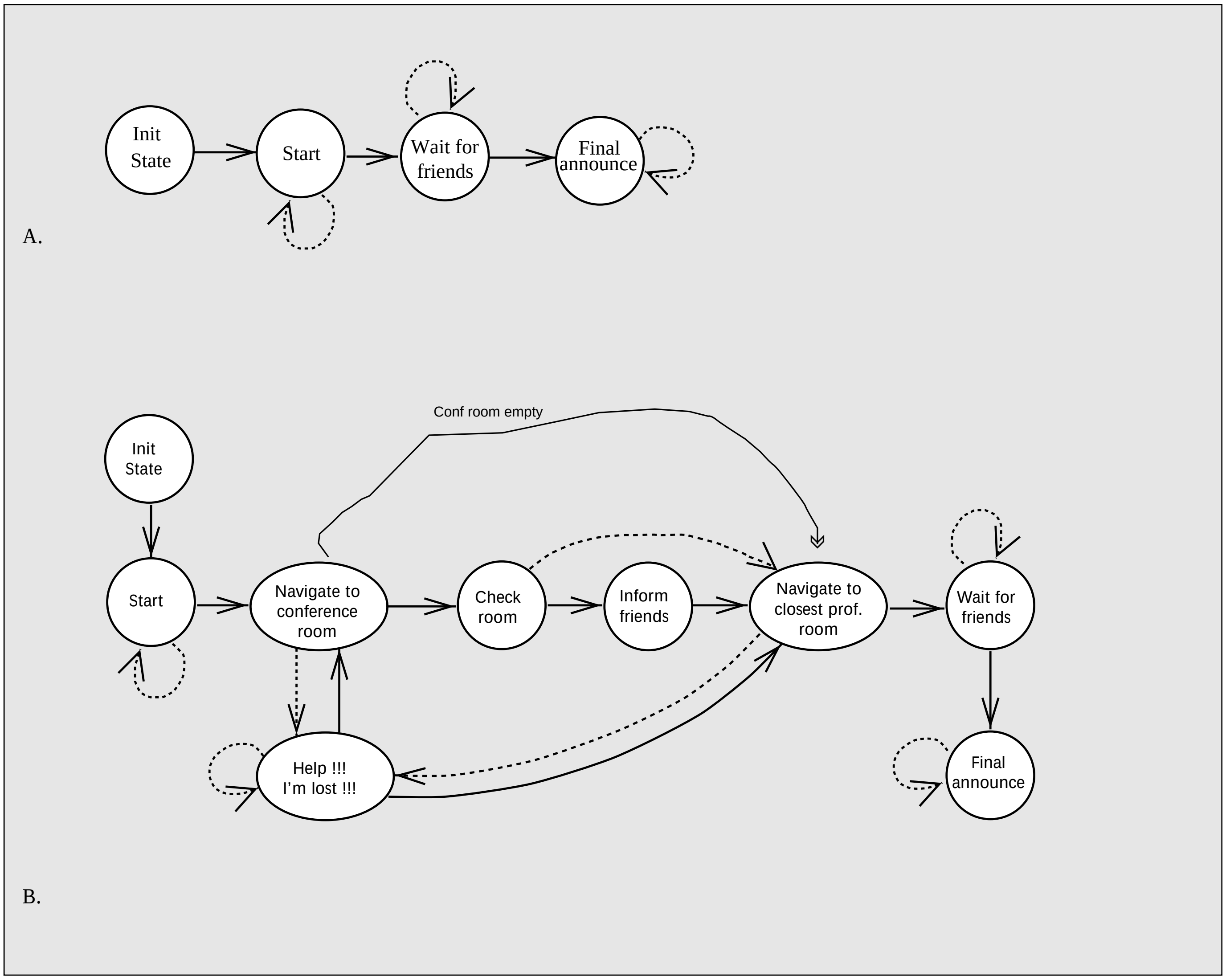


Figure 7. Finite-State Strategy Machines for the Two Types of Robot.
A. Director's office robot. B. Traveling robot.

Given that our map-manager program is an agent, the speech-recognition agent can also be used in the system. Therefore, the user can talk to the system to control the robots or the display. For example, it is possible to say "show me the director's room" to put the focus on this specific room or "robot one, stop, robot one, start" to control a given robot.

Using the global knowledge stored in the database, this application can also generate plans for the robots to execute. The program can be asked (by either a user or a distant agent) to compute the shortest path between two locations, build the corresponding plan, and send it to the robot agent. Plans are locally executed through SAPHIRA in the robots themselves. SAPHIRA returns a success or failure message when it finishes executing the plan; so, the database agent can keep track of the state of all robots. In figure 6, the plan is indicated by a line drawn from the robot to the goal point.

The Strategy Agent

The *strategy agent* controls the coordinated movements of the robots by tracking the total world stated and deciding what tasks each robot should perform at any given moment. Although it would be nice to automatically derive multiagent strategies from a description of the task, environment, and robots, we have not yet built an adequate theory for generating efficient plans. Instead, we built a strategy for the event by hand, taking into account the various contingencies that could arise. The strategy was written as a set of coupled finite-state machines, one for each robot agent. Because the two Pioneer robots had similar tasks, their finite-state machines were equivalent. Figure 7 shows the strategies for these agents.

Note that the finite-state strategies are executed by the strategy agent, not the robots. Each node in the finite-state graph represents a task that the strategy agent dispatches to a robot, for example, navigating to a particular location.

The robot in the director's room has a simple task: Just wait until all the other robots have completed their tasks, then announce the time and place of the meeting. Transitions between states are triggered by events that come into the database: a robot successfully completing a task or some condition becoming known, for example, whether a conference room is empty or full. The dark arrows indicate the successful completion of a task, and the dotted arrows indicate failure.

Both traveling robots have the same strategy: After initialization, they go to a conference room (the strategy agent makes sure they pick different rooms). At any point during navigation, if they get lost, they signal the strategy agent that the navigation was unsuccessful, and the strategy agent asks the mapping agent to return a new location for the robot. This signaling happened several times during preliminary runs when one of the robots attempted to navigate through a conference room and got lost. We were able to tell the robot where it was and keep going from that point.

Arriving at a conference room, the robot checks if the room is empty. If so, it informs the database and continues on to the nearest professor's room. If one robot is navigating to its conference room, and the strategy agent learns that the other robot has found an empty one, it immediately starts the first robot navigating toward the professor's office. Once at the professor's office, the robots announce the expected time of the meeting based on estimates of how long it will take the last robot to reach its professor's office.

Conclusion

The Fifth Annual AAAI Mobile Robot Competition and Exhibition has come a long way since its inception in 1992. At this point, robot navigation in office environments is becoming increasingly reliable, so that we are able to concentrate on interesting problems of high-level strategy, including efficient management of teams of robots. The integration of software agent tools and a multimodal user interface has proven to be a great benefit, making human communication with the robot much more natural and easier to develop. However, there is still a need to develop efficient multiagent planning tools, that is, planners that develop efficient strategies for teams of robots.

References

Brooks, R. A. 1986. A Layered Intelligent Control System for a Mobile Robot. In Proceedings of the IEEE Conference on Robotics and Automation, 14–23. Washington, D.C.: IEEE Computer Society.

Cheyer, A. J., and Julia, L. 1995. Multimodal Maps: An Agent-Based Approach. Paper presented at the International Conference on Cooperative Multimodal Communication (CMC/95), 24–26 May, Eindhoven, The Netherlands.

Cohen, P. R.; Cheyer, A. J.; Wang, M.; and Baeg, S. C. 1994. An Open Agent Architecture. In Software Agents: Papers from the AAAI 1994 Spring Symposium, 1–8. Technical Report SS-94-03. Menlo Park, Calif.: American Association for Artificial Intelligence.

Congdon, C.; Huber, M.; Kortenkamp, D.; Konolige, K.; Myers, K.; and Saffiotti, A. 1993. CARMEL versus FLAKEY: A Comparison of Two Winners. *AI Magazine* 14(1): 49–57.

Julia, L., and Faure, C. 1995. Pattern Recognition and Beautification for a Pen-Based Interface. In Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'95), 7–11 January, Montreal, Canada.

Kameyama, M.; Kawai, G.; and Arima, I. 1996. A Real-Time System for Summarizing Human-Human Spontaneous Spoken Dialogues. In Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP-96), Philadelphia, Pennsylvania.

Konolige, K. 1995. Operation Manual for the Pioneer Mobile Robot, SRI International, Menlo Park, California.

Konolige, K.; Myers, K.; Ruspini, E.; and Saffiotti, A. 1997. The SAPHIRA Architecture: A Design for Autonomy. *Journal of Experimental and Theoretical AI*. Forthcoming.

Moran, D. B., and Cheyer, A. J. 1995. Intelligent Agent–Based User Interfaces. In Proceedings of the International Workshop on Human Interface Technology 95 (IWHIT'95), 7–10. Aizu-Wakamatsu, Fukushima, Japan: The University of Aizu.

Moran, D. B.; Cheyer, A.; Julia, L.; Martin, D.; and Park, S. K. 1997. Multimodal User Interfaces in the Open-Agent Architecture. Paper presented at the IUI97 Conference Proceedings, January, Orlando, Florida.

Moravec, H. P., and Elfes, A. E. 1985. High-Resolution Maps from Wide-Angle Sonar. In Proceedings of the 1985 IEEE International Conference on Robotics and Automation, 116–121. Washington, D.C.: IEEE Computer Society.

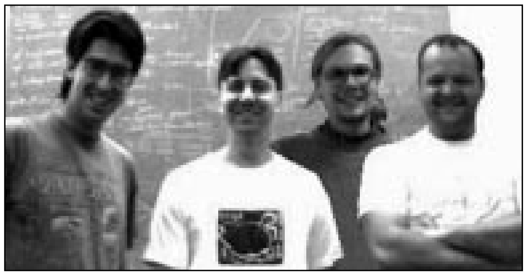
Saffiotti, A.; Konolige, K.; and Ruspini, E. 1995. A Multivalued Logic Approach to Integrating Planning and Control. *Artificial Intelligence* 76(1–2): 481–526.

Didier Guzzoni obtained a bachelor's degree in electrical engineering at the Engineering School of Geneva in 1991 and a Master's degree in computer science at the Swiss Federal Institute of Technology (EPFL) in Lausanne in 1995. After a fellowship in the United States at SRI International, where he worked in the area of agent architectures and mobile robots, he was hired at SFIT as a research assistant to work on a simulator for endoscopic surgery.

Luc Julia is a research engineer in the Speech Technology and Research Laboratory at SRI International. He received his B.S. and M.S. in computer science from Pierre and Marie Curie University, Paris, France, in 1989 and 1990, respectively. He received his Ph.D. from l'Ecole Nationale Superieure des Telecomunications de Paris in 1995. The Ph.D. dissertation gave pattern-recognition methods and solutions to build real multimodal interfaces, integrating speech, gestures, and handwriting. His current research interests include multimodal interaction, speech-enabled web applications, and distributed-agent architectures.

Adam Cheyer is a senior systems programmer in SRI International's Artificial Intelligence Center. He is the chief architect of the open-agent architecture (OAA), a framework for integrating a community of software agents in a distributed environment. Within this framework, Cheyer has authored a number of systems featuring spoken-language and collaborative multimodal (pen and voice) user interfaces to distributed services. He also has expertise in building computer-aided software-engineering tools and expert systems and has published papers in these areas. Cheyer has an M.S. in AI from the University of California at Los Angeles and received the Outstanding Master of Science from the School of Engineering and Applied Science. He received his B.A. with highest honors in computer science from Brandeis University.

Kurt Konolige is a senior computer scientist in the Artificial Intelligence Center at SRI International and a consulting professor at Stanford University. He received his Ph.D. in computer science from Stanford University in 1984; his thesis, "A Deduction Model of Belief and Its Logics," develops a model of belief based on the resource-bounded inferential capabilities of agents. His research interests are broadly based on issues of commonsense reasoning, including introspective reasoning, defeasible reasoning, and reasoning about cognitive state, especially in the context of multiagent systems. More recently, he has conducted research in fuzzy control for reactive systems; constraint-based planning and inference systems; and vision processing and reasoning about perceptual information.



The SRI International Team (l to r): Adam Cheyer, Kurt Konolige, Didier Guzzoni (Swiss Federal Institute of Technology), and Luc Julia.