Interaction With Mobile Robots in Public Places

Sebastian Thrun, Jamie Schulte, Chuck Rosenberg
School of Computer Science
Pittsburgh, PA
{thrun,jscw,chuck}@cs.cmu.edu

1 Introduction

Robotics is undergoing a change. While in the past, robots were predominately employed for assembly and transportation purposes in settings like factory floors, a new generation of service robot has begun to emerge, designed to assist people in everyday life [8, 16, 21, 23]. This raises the question as to what physical shape these robots should take, and how they should behave. In mobile robotics, on which this article focuses, these issues remains relatively poorly explored (despite various efforts, such as [1, 6, 11, 15, 30, 31]), specifically in comparison to mobile robot navigation, which has been investigated thoroughly (see e.g., [3, 5, 14]).

The goal of developing robots that operate in the same spaces where people live and work raises new questions. What physical appearance should these robots have? What things should they by capable of manipulating, and how? What purposes other than manipulation and navigation should such robots have? How should those robots behave? In which way should they interact with people?

This paper focuses on robots operating in public places, where many people interact with a robot for a short time only. Such application domains include receptionists (e.g., in companies, hotels, trade shows), information kiosks, tour-guides, merchandising robots in the retail sector, and so on. The interaction with such robots typically takes place in the *short term*, that is, people spend only limited time with a robot. This critically limits their ability to adapt to the robot. The interaction with robots in public places is also *spontaneous*, which precludes extensive preparations on a person's side (e.g., reading an operation manual). Thus, connecting with modes of interaction people are already familiar with is critical for robots operating in public places.

In the context of spontaneous short-term interaction, we conjecture that certain humanoid features can greatly enhance the effectiveness of a robot. In particular, we believe that human-like physical shape and behavior facilitates the communication from robots to people, and also increases a robot's task effectiveness.

2 Robots In Public Places: A Case Study

2.1 Interactive Museum Tour-Guide Robots

We have recently deployed two autonomous mobile robots operating in public places, with a varying degree of humanoid features. These robots performed the function of interactive tour-guides, deployed in museums to guide people and explain to them what they see along the way [4, 28] (see also [10, 19]). Figure 1 shows pictures of these robots. Rhino, on the left, was the world's first tour-guide robot in a museum: it was deployed for a six-day period in the Summer of 1997 in the Deutsches Museum Bonn

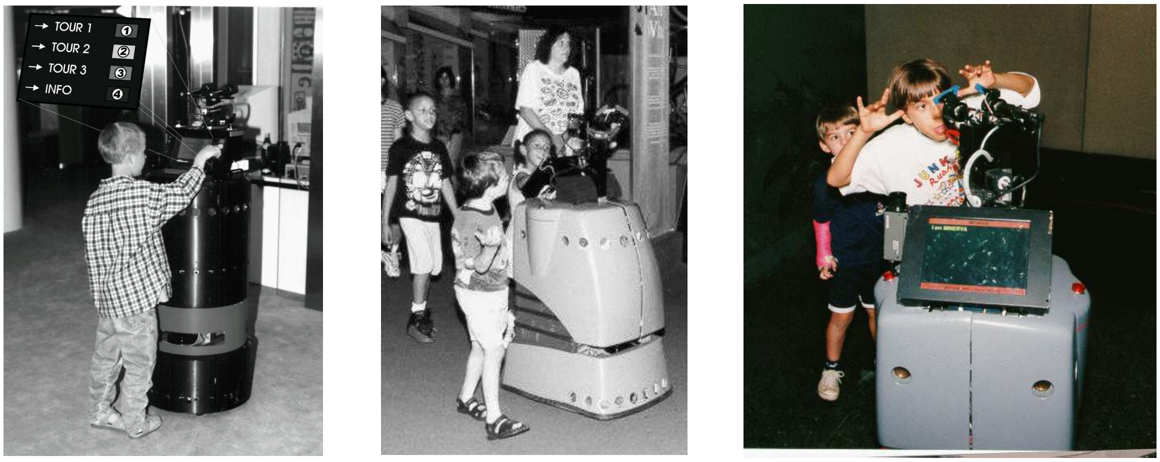

Figure 1: The museum tour-guide robots Rhino (left image) and Minerva (right two images), who autonomously guided thousand of people through crowded museums.

in Germany. Rhino is based on an commercial off-the-shelf B21 robot manufactured by RWI, equipped with a rich sensor suite (cameras, laser, sonar, infrared, tactile). Minerva, a modified RWI B18 robot, was deployed approximately a year later for a period of two weeks in the Smithsonian National Museum of American History in Washington, DC.

The task of tour-guide robots involves three primary goals:

- **Attracting** people to participate in a new tour between tours.
- **Traveling** from one exhibit to the next during the course of a tour.
- **Engaging** people's interest and maintaining their attention while describing a specific exhibit.

Once a person requests a tour, e.g., by touching a robot's touch-sensitive screen, the robot moves from exhibit to exhibit, where it replays pre-recorded acoustic and visual messages. Along the way, people might block the robot's path and otherwise attempt to change the machine's course of action. From the robotics point of view, thus, a key opportunity for interaction arises while traveling from one exhibit to another. Rhino honked a horn to ask for space, whereas Minerva employed a more elaborate strategy to make people aware of its intentions.

One of the most interesting aspects of these robots is actually the way in which they differ. Rhino was primarily developed to test a software system for autonomous navigation crowded environments [4, 29], using natural landmarks for the robot's orientation. Human robot interaction was of secondary importance only. Minerva, on the other hand, was specifically developed with the goal of enhancing people interaction. Minerva basically inherited Rhino's navigation system (with a few extensions), but in addition employed a few, modest humanoid features which we believe made a huge difference in people's perception and, consequently, the effectiveness of the robot.

The need to employ humanoid features arises from the very nature of the domain. Unlike many other service robot applications, most of which involve a long-term interaction of a robot with a specific person [8, 16, 21, 23], the interaction in public places like museums is *spontaneous* and *short-term*. People commonly lack experience in interacting with a robot. In our experience, the typical interaction with a tour-guide robot lasts for less than fifteen minutes, during which a person has to fully grasp the concept and understand how to make use of the machine. Moreover, a tour-guide robot must be self-explanatory, that is, people cannot be required to read instructions before interacting with the robot. It is therefore essential to build robots that connect to what people are readily familiar with [24], hence the (rather modest) use of humanoid features.

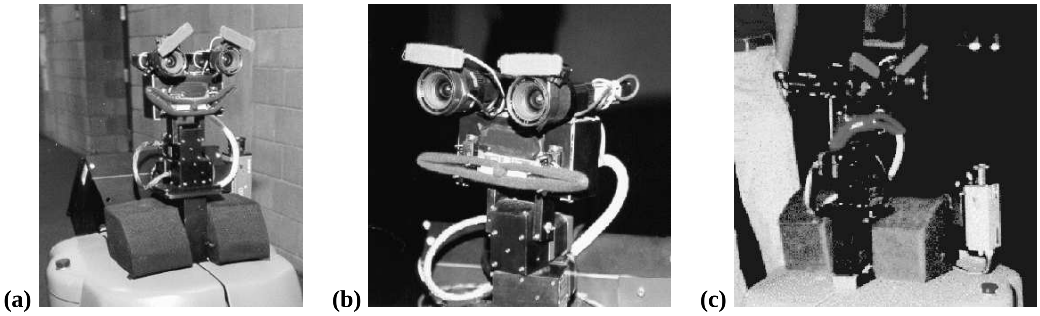

Figure 2: Minerva's face with (a) happy, (b) neutral, and (c) angry facial expressions.

2.2 The Face

Figure 2 shows three possible expressions realized by different configurations of Minvera's face. This face possesses 2 DOFs, the angle of the mouth and the eye-brows. It is mounted on a 2DOF pan/tilt unit, enabling Minerva to face people without physically turning the torso around.

Obviously, the face is only a caricature of a human face [12, 13, 26], which contains specific features related to the expression of simple emotions, while lacking many others. To engage museum visitors, we sought to present as recognizable and intuitive an interface as possible. The face only contains those elements necessary for the degree of expression appropriate for a tour-guide robot. A fixed mask would be incapable of visually representing mood, while a highly accurate simulation of a human face would contain numerous distracting details beyond our control. A physical face was deemed more appropriate than a simulated one displayed on a computer screen, since it can be viewed from arbitrary angles (even from the back), allowing museum visitors to see it without standing directly in front of the robot. As documented in Figure 2, an iconographic face consisting of two eyes with eyebrows and a mouth is almost universally recognizable and can portray the range of simple emotions useful for tour-guide interaction. The face also serves as a *focal point* for the interaction with people.

2.3 Emotional State

Minerva's emotional state is the basis of its travel-related interaction. Travel occurs between stops in a tour when Minerva moves through the museum and finds its way to the next exhibit to discuss. To navigate through crowded spaces, the robot decides whether an obstacle is a human by using a filter applied to the laser range data and the museum map [4]. If the robot is being blocked by a person, it communicates its *intent* to those who are in the way.

Minerva's facial expression is controlled through a finite state machine, where state is represented externally as a mood [9, 17, 20]. This "emotional" state machine encodes the complete travel interaction behavior in a total of four states, as shown in Figure 3. Minerva starts in a "happy" state, smiling while traveling between tour stops, until first confronted by a human obstacle that cannot be trivially bypassed. At this point, the robot kindly announces that it is giving a tour and changes from a smiling to neutral expression, while pointing its head in the direction it needs to travel. If this does not bring success, Minerva adopts a sad expression, and asks the obstructing person to stand behind it. If the person still does not move, then Minerva frowns and becomes even more demanding.

Emotional state helps Minerva achieve navigation goals by enhancing the robot's believability [2, 7]. Observation of interaction with museum visitors suggests that people are generally unconcerned about blocking the path of a passive, mute robot. A change of facial expression and sudden utterance by Minerva usually results in a quick response from anyone in the way. Our subjective interpretation of the

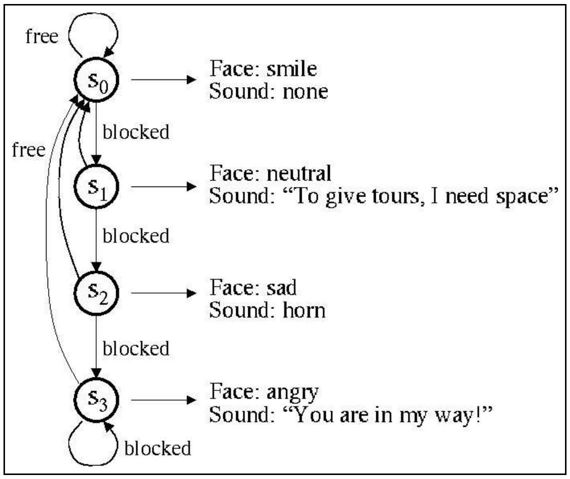

Figure 3: State diagram of Minerva's emotions during travel. "Free" and "blocked" indicate whether a person stands in the robot's path.

effect of emotional state is that the increasing "frustration" of the robot produces feelings of empathy in many people and coerces them to move. This empathy is possible because the timely and exaggerated transition of moods lends Minerva a personality in this limited context.

It is difficult to measure the effectiveness of such an interface quantitatively. In the museum environment, a tour-guide robot is often surrounded by people which impede its forward progress. An examination of the average speed of Minerva (38.8 cm/s) showed it to navigate more quickly than the Rhino robot (33.8 cm/s), even though Minerva operated in a considerably more populated environment [4, 28]. We attribute this to the fact that Minerva could more efficiently and clearly indicate its intended direction of travel. Also, in terms of entertainment value, Minerva's behavior during this time is more interesting to the people who follow the robot. Others have also found interfaces similar to Minerva's to have entertainment value [17, 20].

From observation, it was clear that museum visitors understood the changes in mood brought about by obstructing Minerva. While not everyone chose to move, the robot's *intentions* were quite clear. In the case of the faceless robot Rhino, a horn sound was used to clear people from its path obstructed. People often found this signal to be more ambiguous.

2.4 Learning To Attract People

Another experiment pursued in the museum sought to investigate the feasibility of *learning* patterns of interaction, using people's behavior as a feedback mechanism. Interaction with humans by a robot presents a unique and challenging learning problem. The realm of possible actions with different meanings in an interaction setting is enormous. Subtle changes in the speech timing and volume, or in the intensity of a facial expression can affect the quality of interaction significantly.

For Minerva, the learning only takes place between tours, where the robot spends approximately one minute generating interaction behaviors. At that phase of the interaction, the goal is to attract as many people as possible for the next tour.

Minerva's learning algorithm is a version of *reinforcement learning* [25] with undelayed reward, using *memory-based learning* [18, 22] for generalization. The state comprises the distance and density of people standing near the robot, as measured by the robot's laser range finder. Actions during interaction, are defined to be joint settings of three features: a facial expression, a pan/tilt target for pointing the

Feature	Values
facial expression	happy, neutral, sad, angry
face pointing target	closest person, center of mass of people, least populated area, random direction
sound output	happy speech, “clap your hands”, neutral speech, horn, aggressive speech

Table 1: During the adaptive behavior, an action is performed by setting each of the three features to one of the pre-defined values listed above.

face, and a sound type. The range of possible robot actions was selected to include obviously “good” and “bad” actions, and the overall cadence of interaction and selection of spoken phrases was fixed. The action space for this phase of interaction is outlined in Table 1. Each action persists for 10 seconds, after which a new action is selected. During this interval, the distances and densities of people around the robot are monitored and used to evaluate the effect of the action. The reward function measures the change of the density and proximity of people during those 10 seconds, assuming that people’s behavior is at least partially a result of the robot’s present action.

The memory-based learner stores each individual triplet consisting of a state, an action, and the resulting reward. During learning, a conservative exploration rule was executed similar to pursuit method algorithms known from the k-arm bandit literature [27]. This exploration method, favors actions that performed well in the past, but occasionally selects other actions at random to improve the overall quality of the prediction.

Minerva performed 201 attraction interaction experiments through the two-week deployment, and over time become a more “friendly” robot that attracted people more successfully. Figure 4 shows the learned expected reward for different types of behavior at the end of the experiments. The first plot compares “negative” and “positive” actions. Negative actions are those for which Minerva makes a demand of the visitors in a stern voice while frowning. Positive actions consist of friendlier comments and a neutral or happy facial expression. The larger confidence interval for “negative” actions reflects the fact that less data was collected by Minerva in this less promising region of the action space, since the exploration strategy was biased toward successful actions.

The second plot in Figure 4 compares the expected reward resulting from the five categories of sound that Minerva can produce. Here we can see a clear tendency for happy sounds to produce greater reward than neutral sounds, and for upset sounds to result in a penalty. The fact that the horn sound falls in the neutral reward category sheds some light on the difficulty that Rhino had convincing people to move in previous research. While these figures are of limited significance due to the extremely limited and noisy data, there appears to be a promising trend of increasing reward with friendlier behavior.

3 Summary and Conclusions

In this paper, we have described the interactive component of an interactive museum tour-guide robot called Minerva, which we recently deployed in a museum in the US, and compared it with a robot called Rhino that was deployed previously in a German museum. Both robots perform the same task, and with minor modifications employ the same software for navigation. The more advanced of these robots, the museum tour-guide Minerva, possesses a modest collection of simple, humanoid features that enables

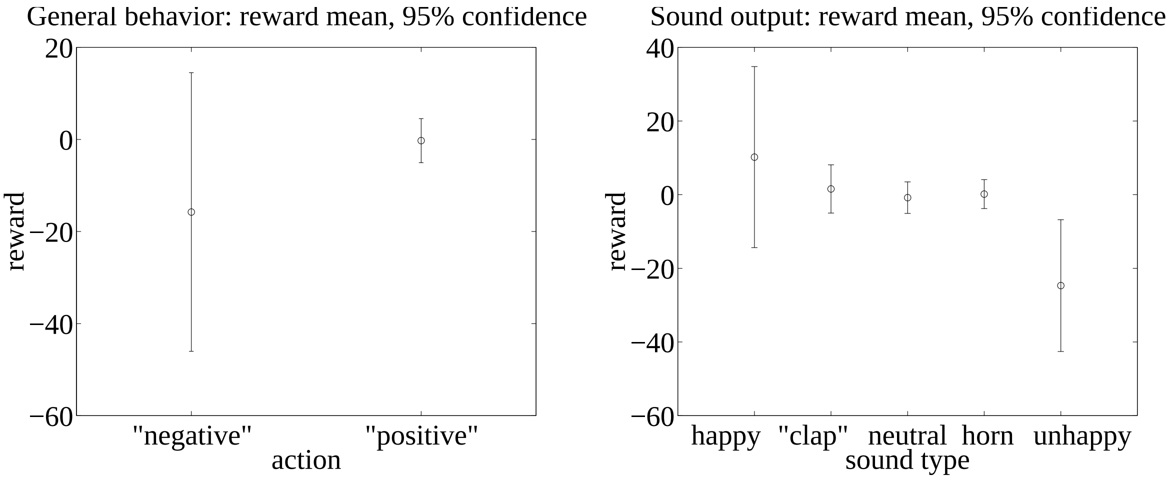

Figure 4: Minerva's expected reward for different sets of actions. *Left*, A comparison of "positive" (friendly) and "negative" (unfriendly) actions. *Right*, five different categories of sounds produced by Minerva, with reward averaged over all other action features.

people who interact with the robot to relate to social aspects they are readily familiar with. This includes an expressive face, a simple emulation of "moods," and a learning algorithm for shaping the robot's interaction on-the-fly.

Unfortunately, the real-world setting did not allow for systematic, prolonged experiments to evaluate the effect of the individual interactive components on the robot's performance. Nevertheless, our experience suggests that human-like features can enhance a robot's effectiveness when interacting with people, specifically in the context of spontaneous short-term interaction. The robot Minerva was able to make progress through the museum during tours at the same rate as the Rhino robot, even though it encountered an order of magnitude more people. Our learning experiments suggested that the robot's behavior had a measurable impact on the degree to which people approached the robot. We also observed that the humanoid features—specifically the motorized face and the finite state automoton emulating simple "emotions"—increased the believability of the robot and enabled people to understand the robot's intentions.

What are the implications of these experiences for service robot design? In robotics, humanoid elements can be grouped into two separate categories:

- "Physical" features that emulate human locomotion and manipulation (e.g., legs, arms), and
- "social" features aimed to facilitate the interaction of robots with people (those might be physical, too, but their main purpose is a social one).

Clearly, our research addresses the second type of humanoid features, hence implications can only be drawn with respect to the social aspects of humanoid robotics. Within the context of social interaction, it has been our strategy *not* to emulate people directly when designing robots. For example, our robots are not equipped with a human-like heads, faces, arms, torsos, etc. Instead, they employ a selected number of vastly simplified icons of human robot interaction, but otherwise differ distinctly from humans. Unfortunately, lack of experience makes us unable to comment on which strategy is more effective: emulating people as closely as technically possible, or instead developing robots that differ distinctly but possess a small number of carefully chosen humanoid features. Nevertheless, our experience suggests that humanoid features can greatly enhance the interaction of robots with people, thereby increasing a robot's effectiveness and user acceptance. These conclusions apply specifically in the context of spontaneous short-term interaction, hence are particularly relevant for robots that operate in public places such as shopping malls, trade shows, hotels, amusement parks, and airports, where people's ability to adapt to non-intuitive interfaces is generally extremely limited.

Acknowledgments

Aspects of this research have been sponsored by the National Science Foundation (and CAREER grant number IIS-9876136 and regular grant number IIS-9877033), and by DARPA-ATO via TACOM (contract number DAAE07-98-C-L032) and DARPA-ISO via Rome Labs (contract number F30602-98-2-0137), which is also gratefully acknowledged. The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing official policies or endorsements, either expressed or implied, of the United States Government.

References

[1] H. Asoh, S. Hayamizu, H. Isao, Y. Motomura, S. Akaho, and T. Matsui. Socially embedded learning of office-conversant robot jijo-2. In *Proceedings of IJCAI-97*. IJCAI, Inc., 1997.

[2] J. Bates. The role of emotion in believable agents. Technical Report CMU-CS-94-136, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, April 1994.

[3] J. Borenstein, B. Everett, and L. Feng. *Navigating Mobile Robots: Systems and Techniques*. A. K. Peters, Ltd., Wellesley, MA, 1996.

[4] W. Burgard, A.B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2):3–55, 1999.

[5] I.J. Cox and G.T. Wilfong, editors. *Autonomous Robot Vehicles*. Springer Verlag, 1990.

[6] J. Crisman and G. Bekey. Grand challenges for robotics and automation: The 1996 ICRA panel discussion. ICRA-96, 1996.

[7] K. Dautenhahn. The role of interactive conceptions of intelligence and life in cognitive technology. In *Proceedings of CT-97*, pages 33–43, 1997.

[8] G. Engelberger. Services. In Shimon Y. Nof, editor, *Handbook of Industrial Robotics*, chapter 64., pages 1201–1212. John Wiley and Sons, 2nd edition, 1999.

[9] C. Breazeal (Ferrell). A motivational system for regulating human-robot interaction. In *Proceedings of AAAI-98*, pages 54–61, Madison, WI, 1998.

[10] I. Horswill. Polly: A vision-based artificial agent. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93)*. MIT Press, 1993.

[11] R.E. Kahn, M.J. Swain, P.N. Prokopowicz, and R.J. Firby. Gesture recognition using the perseus architecture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 734–741, San Francisco, CA, 1996.

[12] W. King and J. Ohya. The representation of agents: Anthropomorphism, agency, and intelligence. In *Proceedings of CHI-96*, 1996.

[13] T. Koda. Agents with faces: The effect of personification. In *5th IEEE International Workshop on Robot and Human Communication*, Tsukuba, Japan, November 1996.

[14] D. Kortenkamp, R.P. Bonasso, and R. Murphy, editors. *AI-based Mobile Robots: Case studies of successful robot systems*, Cambridge, MA, 1998. MIT Press.

[15] D. Kortenkamp, E. Huber, and P. Bonasso. Recognizing and interpreting gestures on a mobile robot. In *Proceedings of AAAI-96*, pages 915–921. AAAI Press/The MIT Press, 1996.

[16] G. Lacey and K.M. Dawson-Howe. The application of robotics to a mobility aid for the elderly blind. *Robotics and Autonomous Systems*, 23:245–252, 1998.

[17] P. Maes. Artificial life meets entertainment: Interacting with lifelike autonomous agents. *Special Issue on New Horizons of Commercial and Industrial AI, Communications of the ACM*, 38(11):108–114, November 1995.

[18] A. W. Moore. *Efficient Memory-based Learning for Robot Control*. PhD thesis, Trinity Hall, University of Cambridge, England, 1990.

[19] I.R. Nourbakhsh. The failures of a self-reliant tour robot with no planner. Can be obtained at `http://www.cs.cmu.edu/∼illah/SAGE/index.html`, 1998.

[20] C. Rosenberg and C. Angle. IT: An interactive animatronic prototype. IS Robotics Inc. internal development project, description available at: `http://www.cs.cmu.edu/∼chuck/robotpg/itpg/`, December 1994.

[21] N. Roy, G. Baltus, D. Fox, F. Gemperle, J. Goetz, T. Hirsch, D. Magaritis, M. Montemerlo, J. Pineau, J. Schulte, and S. Thrun. Towards personal service robots for the elderly. In *Proceedings of the Workshop on Interactive Robotics and Entertainment (WIRE)*, Pittsburgh, PA, 2000. Carnegie Mellon University.

[22] S. Schaal and C. G. Atkeson. Assessing the quality of learned local models. In J.D. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems 6*, San Mateo, CA, 1994. Morgan Kaufmann.

[23] R.D. Schraft and G. Schmierer. *Serviceroboter*. Springer verlag, 1998. In German.

[24] J. Schulte, C. Rosenberg, and S. Thrun. Spontaneous short-term interaction with mobile robots in public places. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1999.

[25] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.

[26] A. Takeuchi and T. Naito. Situated facial displays: Towards social interaction. In *Proceedings of CHI-95*, 1995.

[27] M. A. L. Thathachar and P. S. Sastry. Estimator algorithms for learning automata. In *Proceedings of the Platinum Jubilee Conference on Systems and Signal Processing*, Bengalore, India, 1986.

[28] S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. MINERVA: A second generation mobile tour-guide robot. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1999.

[29] S. Thrun, A. Bücken, W. Burgard, D. Fox, T. Fröhlinghaus, D. Henning, T. Hofmann, M. Krell, and T. Schmidt. Map learning and high-speed navigation in RHINO. In D. Kortenkamp, R.P. Bonasso, and R Murphy, editors, *AI-based Mobile Robots: Case Studies of Successful Robot Systems*. MIT Press, 1998.

[30] M. C. Torrance. Natural communication with robots. Master's thesis, MIT Department of Electrical Engineering and Computer Science, Cambridge, MA, January 1994.

[31] S. Waldherr, S. Thrun, and R. Romero. A gesture-based interface for human-robot interaction. *Autonomous Robots*, 9(2), 2000. to appear.