Designing Robots for Long-Term Social Interaction*

Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell,
Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C. Schultz[†], and Jue Wang

Carnegie Mellon University, Pittsburgh PA
[†]*Naval Research Laboratory, Washington DC*

Abstract— **Valerie the Roboceptionist is the most recent addition to Carnegie Mellon's Social Robots Project. A permanent installation in the entranceway to Newell-Simon Hall, the robot combines useful functionality—giving directions, looking up weather forecasts, etc.—with an interesting and compelling character. We are using Valerie to investigate human-robot social interaction, especially long-term human-robot "relationships." Over a nine-month period, we have found that many visitors continue to interact with the robot on a daily basis, but that few of the individual interactions last for more than 30 seconds. Our analysis of the data has indicated several design decisions that should facilitate more natural human-robot interactions.**

Index Terms— **robotics, social robots, human-robot interaction**

I. INTRODUCTION

While many researchers are investigating human-robot social interaction, one area that remains relatively unexplored is that of continued long-term interaction. The Roboceptionist ("robot receptionist") Project, part of the Social Robots Project, is investigating how a social robot can remain compelling over a long period of time—days, weeks, and even years.

Our approach is to create a robot that can provide useful services, but that also exhibits personality and character. The robot was designed for ease of interaction without requiring any training or expertise, and to be compelling enough to encourage multiple visits over extended periods of time.

The character we have designed, named Valerie, is built from a mobile base with a moving flat-panel monitor mounted on top, which displays a graphical human-like face. Valerie remains stationary inside a small booth near the main entrance of Newell-Simon Hall at Carnegie Mellon University (Fig 1). Anyone who walks through the building, including students, faculty, and visitors, can interact with the robot.

II. BACKGROUND

The Social Robots Project began with the goal of investigating human-robot social interaction. Experiments with the robot Vikia studied the effects of attentive movement and an animated face on people's willingness to engage in a short interaction with a robot [1]. These experiments confirmed the group's intuitions that both movement and a recognizable face have a positive impact on human-robot social interaction. Grace, a joint project by our group and a number of other research institutions, has participated in the AAAI robot challenge for several years [2]. The challenge requires a robot to

Fig. 1.   Valerie the Roboceptionist, in her booth.

register for the conference, find the room it is scheduled to speak in, and give a short talk about its own capabilities. Social interaction is vital to performing these tasks successfully. Grace uses conversational capabilities similar to Valerie's to interact with workers at the registration desk in a socially appropriate manner.

A number of other research groups are also using robots to explore social interaction. Kismet [3] and Sparky [4] both used facial expression and movement to interact with humans. Unlike Valerie, these robots engaged in only short-term, non-verbal interactions, and their purpose was not to provide users with useful information. On the other hand, a number of robots have been designed over the years to serve as tour guides for museum visitors [5]–[7]. Like Valerie, their purpose is to inform as well as to entertain. These robots also use speech capabilities to provide users with useful information, and they use facial and emotional expressions to improve the quality of interaction. However, these interactions are fairly structured and primarily one-way—people do not actively converse with the robots. The Nursebot [8] is another robot that uses social competence to improve task performance. That project's goals were similar to our own in that it aimed to create a robot that

engaged in repeated interactions with people over an extended period of time. Robovie, an interactive humanoid robot, has been used in long-term interaction studies with children, but its designers noted that it "failed to keep most of the children's interest after the 1st week" [9]. With Valerie, we hope to maintain interest over longer periods.

III. DESIGN DECISIONS

The Roboceptionist Project is the product of a collaboration between the Robotics Institute and the School of Drama at Carnegie Mellon. Planning and design was conducted for almost a year prior to Valerie's deployment. Some of the major design decisions are detailed below.

A. A Receptionist

We wanted the robot to be familiar and non-threatening to people who access the building (primarily non-roboticists). We chose a receptionist as Valerie's role for several reasons:

- Receptionists have frequent interaction with the public, and people have well-understood expectations for how to interact with receptionists.
- Valerie is capable of handling some of the tasks that a receptionist would perform, such as looking up office numbers and providing directions.
- We could station the robot in a public space in order to maximize the number of interactions with humans. In addition, the robot could be located behind a desk, which provides some security for the hardware.

B. Character and Personality

In order to make the robot a compelling presence, we elected to make it human-like in its interactions. The Drama group helped to imbue the robot with human characteristics by giving her a name, a personality, a back-story, and several storylines that unfold over time. Events in her life are related in "conversation" to visitors who stop to chat with her. In addition, people can keep up with Valerie's life online at http://www.roboceptionist.com.

Valerie enables a new form of storytelling. Her entire story, as well as character-related vocalizations and behaviors, were scripted by students in the School of Drama. Complex story-lines interweave and evolve over a period of several months. For example, she has an active love life and a singing career that she pursues in her free time. Writers and designers must deal with a character that has no vocal intonation, no natural facial expressions, and no form of natural movement. Funda-mental assumptions regarding the creation of live storytelling had to be reviewed; what works with humans often does not work with robots.

C. The Graphical Face

Valerie's "head" is a flat-screen LCD monitor mounted on pan-tilt unit. Her "face," shown in Fig. 2, is a graphically rendered 3D model. Her facial modeling and expressions were created by members of the Drama group. Choosing a graphical, rather than mechanical, face was a significant design

Fig. 2. The graphical face.

decision. The flat-screen face offers several advantages over a mechanical face:

- The graphical face is very expressive, with the ability to move individual muscles to generate a wide range of facial expressions.
- A mechanical face is less reliable than a graphical one, due to its many moving parts.
- Changes can easily be made to the graphical face. For example, as part of one story, Valerie's hairstyle changed. A physical mechanism would be more difficult to modify.

The greatest disadvantage of the graphical face is that it lacks the physical embodiment of a mechanical face. In particular, although the head rotates to face visitors, it can be difficult to determine exactly where the robot is looking.

D. Interaction Structure

Decisions about the mode and structure of interactions were driven by a desire to ensure that visitors do not become frus-trated with the system and are satisfied with the interactions.

1) Storytelling: One of Valerie's primary interaction modes is storytelling. Valerie's story is told in a very human way: subjectively and evolving over time. Her story is revealed through monologues, which are styled as phone conversations with characters in her life. The writers from Drama crafted four storylines that evolved over the school year: Valerie's social life, her lounge singing career, her therapy business, and her job as a receptionist. Storytelling was chosen in order to make the robot appear more human-like and thus to allow visitors to interact easily with her.

Valerie's evolving life stories follow a well-known model— that of the soap opera, or of the currently popular "reality" show. By making Valerie a compelling character, we hoped to encourage people to visit the robot repeatedly over time, in the same way that they might eagerly tune in to each episode of *As the World Turns* or *Survivor*.

2) Keyboard Input: Both speech and keyboard input modal-ities were considered for visitors' interactions with Valerie. Speech is more natural for most people, but keyboard input is easier to control and more reliable than any general speech recognition systems, which typically require either training for individual users or a drastic reduction in the allowable vocabulary [10]. In addition, having visitors interact vocally was unlikely to be robust due to the placement of the robot in

a busy hallway. Speech recognition systems are generally poor at handling noise and echoes in the environment. While a head-worn microphone can reduce the effect of ambient noise, we felt that requiring visitors to first don a headset would detract significantly from the overall experience.

3) Chatbot: For handling natural language input, the decision to use a pattern-matching "chatbot" rather than a grammatical parser was based on the ease of adding information and being be able to recognize novel sentences. Grammatical parsing would make new sentences difficult to add, requiring additions to the dictionary and to the grammar, and few existing systems can handle sentence fragments well.

The rule-based pattern-matching system that was chosen for Valerie is modified from Aine [11], which is in turn derived from AIML and ALICE [12]. The rules are simple to write, can return any type of desired data (including tags usable by other components), and allow many different wordings of sentences to be recognized with just a few rules. Unlike a robust parser, Aine is completely ignorant of the language—knowing only the order of words—so it cannot take advantage of context for better recognition. On the other hand, it can handle ungrammatical sentences, sentence fragments, and even many misspellings, if appropriate rules have been added to the Aine database.

IV. ROBOT CAPABILITIES

Valerie's body is a B21r mobile robot produced by iRobot. Valerie has an expressive computer-animated face (Fig. 2) that is displayed on a 15" flat-panel LCD screen mounted on a pan-tilt unit. By panning the display, Valerie is able to "look" as much as 120 degrees to either side.

The animated face is one of the most important aspects of Valerie's ability to interact with humans. It is used both for emotional expression and for simple head gestures, since Valerie lacks any conventional manipulators. The face is based on an implementation of the simple face in [13]. It incorporates a muscle-level model of face movement to allow semi-realistic face motions.

Valerie's speech is generated from text using the Theta text-to-speech engine developed by Cepstral, and is combined with automatic lip-syncing of the face musculature. The text is also displayed in "balloons" next to the face (Fig. 2), to aid human comprehension of the synthetic voice. Additional facial movements and expressions can be synchronized with speech, using a graphical tool developed in-house. The tool allows developers to move the facial muscles interactively to generate facial expressions, and to script these expressions along a timeline, synchronized with the speech.

A SICK scanning laser range finder is used to detect and track individuals as they move through the space surrounding the robot's work area. The range finder is located in the booth behind a slit in the front wall, providing a 180-degree field of view at roughly thigh-height. To estimate people's locations, the software filters out range readings that correspond to the (known) background and then clusters and tracks (via Kalman filtering) the remaining range readings.

Fig. 3.   The user interface.

Valerie's user interface consists of a keyboard, monitor, and magnetic card-stripe reader (Fig. 3). The monitor displays a graphical user interface that echoes the user's keystrokes, stores a brief history of the inputs, and provides feedback about whether a card run through the cardreader was successfully read. The cardreader is mounted between the keyboard and user interface display and allows a visitor to swipe any magnetic-stripe card in order to uniquely identify himself (e.g. Carnegie Mellon ID cards or many drivers' licenses). The data from the card is stored in a one-way MD5 hash and is used as a key to remember pertinent user information from one interaction to the next.

V. INTERACTION SETUP

Natural human conversation goes through several phases, from greeting through departure. Greeting is key to initiating interaction [14], so Valerie tries to greet those people near the booth who might be interested in engaging in conversation. The area surrounding the booth covered by the laser range finder is divided into several regions that are used for classifying detected people into "attentional" states (see Fig. 4). People who are several feet away or moving quickly past the booth are classified as "present," and Valerie pays no attention to them. Individuals who are closer are classified as "attending," and Valerie greets these people by turning to them and vocalizing if she is otherwise idle. Visitors who are next to the desk but off to the side are classified as "engaged," and Valerie acknowledges their presence but does not expect input from them; they may, for example, be observing an interaction under way. Finally, those who are directly in front of the keyboard are considered to be "interacting," and Valerie repeatedly prompts them for input if they are not typing.

The text input interface was designed to be clear and simple to use, to allow for ease of interaction. Valerie senses whenever a person begins typing, and presents an attentive expression, often accompanied with a friendly, "How may I help you?" In addition, Valerie will prompt a visitor if he does not complete his input by pressing "Enter." When a person leaves the "interacting" region, Valerie signals the end of the interaction by saying "goodbye."
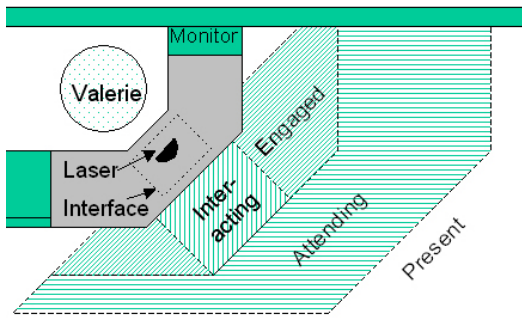
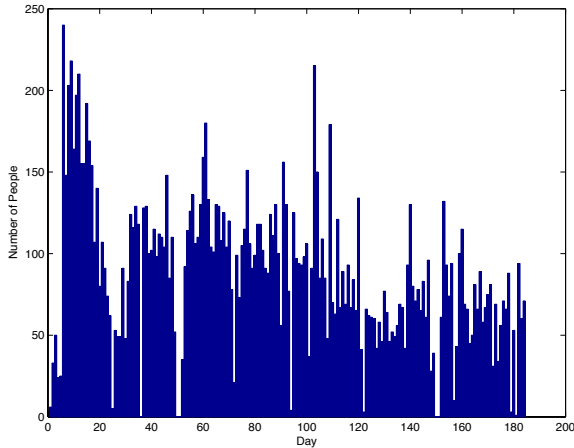Fig. 4. The attentional zones surrounding the booth.

Fig. 5. Number of people who interact with Valerie per day. The decrease after day 120 corresponds to the start of the university's summer vacation.

VI. RESULTS

Valerie first opened for general interaction at the end of November 2003. Since then, the robot has been available for approximately eight hours a day, five days a week, excluding holidays and a few days of down time. Here, we report on an analysis of the first nine months of operation, a total of more than 180 days.

A. General Visitors

During those nine months, people have interacted with Valerie—by typing at least one line of text—over 16,000 times. A large number of people interact with Valerie every day, with a daily average of over 88 interactors. Fig. 5 charts the number of people who have interacted with her over the nine months. Days where the number of interactors is 0 generally correspond to university holidays. The average decrease in the daily visitors following day 120 is most likely a result of summer vacation, when fewer people pass through the building. Even during that time period, however, the average number of interactors was still over 60 per day.

Fig. 6 graphs the median time visitors spent interacting with Valerie each week. During the first week of Valerie's appearance in the building, people tended to interact with her for longer periods of time—a few staying for as long as an hour or more, simply testing the robot's capabilities. For this
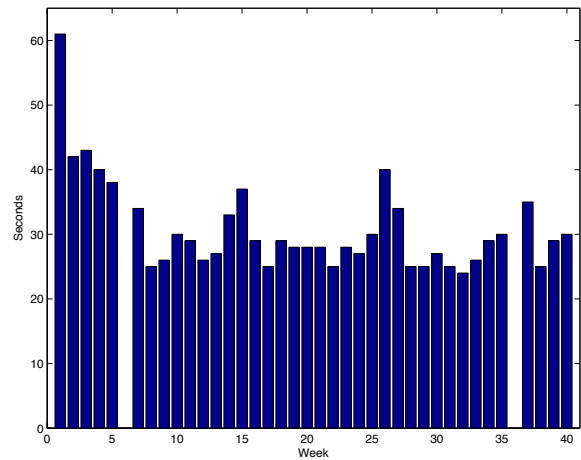

Fig. 6. Median time spent interacting with Valerie per week. A novelty effect is evident during the first few weeks, but interaction times remain consistent afterward.

reason, averages are not reported here; the median values better reflect the "average" visitor. After the initial "novelty" of the robot faded, typical interactions with the robot leveled off at just under half a minute—long enough to ask for directions, or to exchange a few pleasantries.

B. Storytelling

To provide a compelling reason for people to visit the robot repeatedly over time, Valerie's stories were changed every other week. People have to visit repeatedly in order to hear how the storylines evolve. Fig. 7 graphs how many times people listened to any monologue, on a weekly basis. Only people who stayed for at least 20 seconds of a monologue are included. Weeks 1-6 are not shown due to missing data on how long individuals listened to the stories. Note that the university's summer vacation began week 26, resulting in the sudden dropoff. Also, in week 25 all the storylines came to a climax, which perhaps accounts for the spike in listeners that week. Overall, an average of 254 people per week (or about 50 per day) were near the booth for at least 20 seconds while Valerie gave a monologue. People may have heard multiple monologues (and are counted for each story), and often are close enough to the booth to hear a story without ever interacting with the robot.

Valerie's monologues typically lasted 2-3 minutes. However, most visitors did not stay for a significant portion of the stories. To calculate the times that people spent listening to a story, only individuals who entered the "attending" zone and were present for at least five seconds were counted, since that would imply that they at least paused near the robot while she was talking. Other than a few spikes, the weekly median time visitors listened to a monologue hovered around 12 seconds.

C. Repeat Visitors

Currently, Valerie cannot automatically identify her visitors. A visitor must swipe an ID card on each visit to identify himself to the robot. In the first nine months of operation,
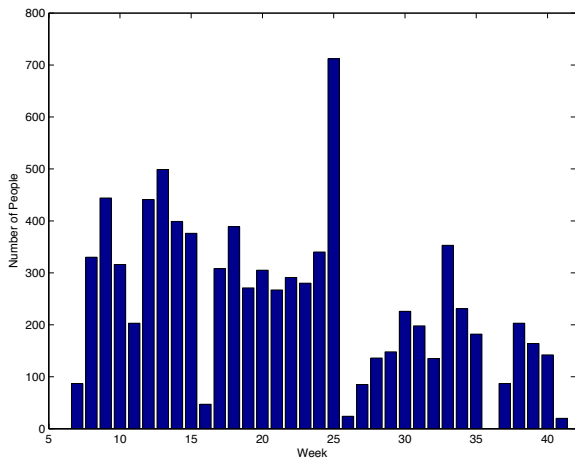
Fig. 7. Counts of people present for at least 20 seconds during one of Valerie's monologues, per week. The dropoff at week 26 correlates to the university's summer vacation beginning.

Fig. 8. The number of times per week that Valerie was visited by a person identified as having visited the robot multiple times.

753 different people have used an ID card in this manner at least once. Of these, only 233 visitors swiped a card during a subsequent visit. However, swiping a card is inconvenient, and currently provides little benefit to the person beyond being greeted by name. Regardless, we have anecdotal evidence of people repeatedly interacting with Valerie, such as greeting her each morning as they enter the building.

Fig. 8 depicts the number of times per week that Valerie was visited by a person who was tracked as visiting the robot on more than one occasion. The cardreader suffered from intermittent hardware failure, which likely accounts for the weeks of few or no visitors. As with the total number of visitors, a novelty effect was present over the first few weeks, but settled over time. Again, the weeks preceding week 26 correspond to university finals and the start of summer vacation. On average, approximately 7 of the people Valerie interacted with each day were identified as repeat visitors. Repeat visitors interacted with the robot for considerably longer periods than other visitors, typically staying for a minute or longer. In addition, repeat visitors typically listened to monologues for longer periods, generally listening for just under a minute.

D. Impact of "Default Response"

Because of the simplicity of Valerie's chatbot system, she often replies to input with one of several default responses that indicate her lack of understanding. Even so, Valerie gave such a response to only 28.1% of all the inputs typed to her. Just over half (52.9%) of the people who interacted with Valerie received a default response at least one time during their interaction. Of those people, fully 74.9% typed at least one more input to the robot, and 57.3% stayed to type at least two more inputs. In other words, most people did not simply walk away from the robot after receiving a generic response from her.

VII. DISCUSSION

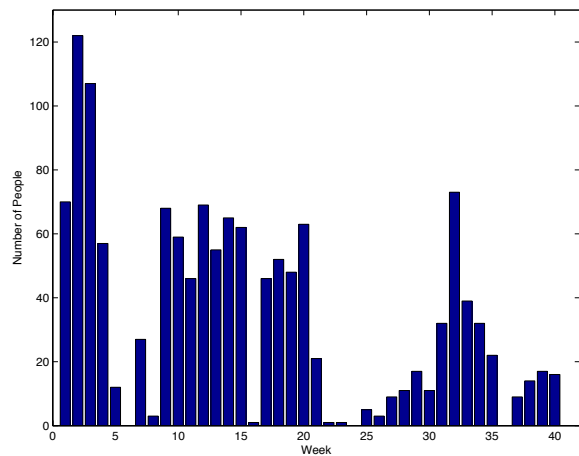Overall, we found that, even after the "novelty effect" faded, Valerie continued to have a steady stream of visitors on a daily basis. In addition, though identifying oneself to the robot is somewhat of a hassle, well over 200 people chose to do so multiple times over the nine-month period. Contrary to our expectation that people would be highly engaged by the stories, most people who interacted with the robot stayed long enough to greet the robot and hear one or two sentences of a monologue, but not more. In contrast, people whom we know were repeat visitors interacted and listened for much longer periods. This makes sense intuitively, since the people who are interested enough in the robot to interact with it multiple times are also likely to want to interact with the robot for longer times.

VIII. FUTURE WORK

Based on this analysis of the interaction data, we are considering several additions and changes that we expect will enhance the interaction experience, leading to longer and more satisfying interactions.

A. New Story-Telling Mechanism

Currently, most people do not stay to listen to all of Valerie's monologues. One way to encourage longer interactions with the robot may be to make the storytelling more interactive. We have implemented, and are currently testing, a new storytelling mechanism that encourages more turn-taking in a dialogue by requiring users to indicate their interest in the story. This, in turn, allows them to explore in more detail aspects of the stories that are of particular interest to them.

B. Person Identification and Personalization

To study true long-term interactions with Valerie, the robot needs a better way to identify repeat visitors. We suspect that the current "card swipe" identification method is too cumbersome, and so are currently investigating ways for Valerie to identify people automatically, through visual face recognition and/or radio frequency identification (RFID) tags. We are also working to obtain a more accurate model of users' attentional

state (present, attending, engaged, etc.) through the use of visual head-pose estimation.

To establish long-term relationships, Valerie should not only identify but also "get to know" people who frequent the booth. Since people who know each other share a common set of beliefs about each other and about their conversation together, establishing this "common ground" is a key element of human-human interactions [15]. If Valerie can learn about a person's interests (such as storylines that they prefer), personalizing the interactions will perhaps make the experience more enjoyable.

C. Emotions

If Valerie is to act human-like, she should respond to events in her life in an emotional manner. For example, Valerie could become visibly happy when greeting a person who interacts with her frequently—or annoyed if that person is often rude. If Valerie becomes frustrated at not understanding a person's questions, then perhaps that person will attempt to rephrase the questions to help the robot to understand. We are developing a model based on the OCC theory of emotions [16] to give Valerie some of this capability.

D. Robust Parser + Aine

With Aine, Valerie has extremely limited language understanding. Combining a robust parser with Aine as a fallback mechanism would likely result in a language system that understands significantly more sentences or sentence fragments than the current system does.

IX. DESIGN RECOMMENDATIONS

We believe that long-term human-robot social interaction can be greatly improved through an understanding of human-human social interaction. While we attempted to incorporate some of the knowledge of human social behavior into Valerie, it remains unclear exactly how much of such knowledge is truly applicable to human-robot interaction. Based on our experience observing and logging the interactions with Valerie, we propose the following design recommendations:

- *Greeting*. Greeting should be used to make the robot engaging, to shape expectations for the ease of interacting with it. The robot should have an approachable interface in order to foster conversational interaction with expert users, novices, and single participants as well as groups.
- *Dialogue*. Turn taking in dialogue should be supported. We learned that with the monologues, few stayed to hear more than 1-2 sentences of the story, which we believe resulted from the lack of interaction with the robot during the storytelling time. The robot should avoid being a "motor-mouth," and instead provide more natural dialogue with people.
- *Engagement*. Allow for "common ground" to be established between the robot and repeat visitors. The structure, length, and language structure of interactions should be differentiated for repeat visitors to encourage them to engage in dialogue again and again.

- *Departure*. Provide a mechanism for ending interaction that is based on human social norms. Valerie would typically continue to tell her stories even after people departed from the area. Instead, the robot should sense the end of a focused interaction with a visitor, and give a goodbye salutation when it occurs.

X. CONCLUSION

Most social robot projects have worked to create systems that recognize and exhibit human emotions, or that aim solely to convey information. However, the range of capabilities exhibited by these robots is typically discovered and exhausted by people rather quickly, and such robots do not maintain their users' interest over the long term. This is problematic for an interactive robot that is situated in an environment for a long time. We propose that endowing such a robot with personality, character, and a story that changes over time will keep people interested enough to provide the robot with a steady stream of visitors. When the present interaction is tied to a past and future narrative, the limitations in the robot's interactive capabilities might be bolstered by its ability to entertain.

Valerie represents an initial attempt to develop a robot that is compelling to interact with over a long period of time. She has continually attracted and engaged many visitors on a daily basis for the past nine months, and continues to do so. We are now working toward making her behavior more human-like, in an effort to improve the quality of her interactions with her visitors. We believe this will aid in making the robot more compelling for long-term interactions.

REFERENCES

[1] A. Bruce, I. Nourbakhsh, and R. Simmons. The role of expressiveness and attention in human-robot interaction. In *IEEE Conference on Robotics and Automation*, 2002.

[2] R. Simmons et al. Grace: An autonomous robot for the AAAI robot challenge. *AAAI Magazine*, 42:2:51–72, Summer 2003.

[3] C. Breazeal. and B. Scassellati. How to build robots that make friends and influence people. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS1999) Kyonju, Korea.*, 1999.

[4] M. Scheeff. Experiences with Sparky: A social robot, 2000.

[5] Wolfram Burgard et al. Experiences with the interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2):3–55, 1999.

[6] T. Willeke, C. Kunz, and I. Nourbakhsh. The history of the mobot museum robot series: An evolutionary study, 2001.

[7] R. Siegwart et al. Robox at Expo.02: A large scale installation of personal robots. *Robotics and Autonomous Systems, Special issue on Socially Interactive Robots*, 42:203–222, 2003.

[8] M. Montemerlo, J. Pineau, N. Roy, S. Thrun, and V. Verma. Experiences with a mobile robotic guide for the elderly, 2002.

[9] Takayuki Kanda, Takayki Hirano, Daniel Eaton, and Hiroshi Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19:61–84, 2004.

[10] R. Cole, J. Mariani, H. Uszkoreit, A. Zaenen, and V. Zue. Survey of the state of the art in human language technology, 1995.

[11] David Calinski. http://www.neodave.civ.pl/aine/.

[12] A.L.I.C.E. http://www.alicebot.org/.

[13] F. Parke and K. Waters. *Computer Facial Animation*. A.K. Peters, Ltd., December 1996.

[14] Adam Kendon and Andrew Ferber. A description of some human greetings. In *Conducting Interaction*. Cambridge University Press, 1990.

[15] Herbert H. Clark. *Using Language*. Cambridge University Press, 1996.

[16] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.