The Humanoid Museum Tour Guide Robotinho

Felix Faber Maren Bennewitz Clemens Eppner Attila Görög
Christoph Gonsior Dominik Joho Michael Schreiber Sven Behnke

Abstract— **Wheeled tour guide robots have already been deployed in various museums or fairs worldwide. A key requirement for successful tour guide robots is to interact with people and to entertain them. Most of the previous tour guide robots, however, focused more on the involved navigation task than on natural interaction with humans. Humanoid robots, on the other hand, offer a great potential for investigating intuitive, multimodal interaction between humans and machines. In this paper, we present our mobile full-body humanoid tour guide robot Robotinho. We provide mechanical and electrical details and cover perception, the integration of multiple modalities for interaction, navigation control, and system integration aspects. The multimodal interaction capabilities of Robotinho have been designed and enhanced according to the questionnaires filled out by the people who interacted with the robot at previous public demonstrations. We present experiences we have made during experiments in which untrained users interacted with the robot.**

Fig. 1. Robotinho guides visitors to exhibits and presents them. It uses multiple modalities to interact with people in a natural way.

I. Introduction

Humanoid robots have become a popular platform in recent years. More and more research groups worldwide develop complex machines with a human-like body plan and human-like senses [1], [2], [3]. One of the most important motivations for many humanoid projects is that such robots could be capable of intuitive, multimodal interaction with people. This makes humanoid robots an ideal platform for museum tour guide projects. The museum application furthermore allows researchers to evaluate their approaches to intuitive, multimodal interaction in the public.

Advanced tour guide robots should make use of multiple modalities such as speech, facial expressions, gestures, body language, etc. to interact with people. If successful, this approach yields a user interface that leverages the evolution of human communication and that is intuitive to naive users, as they have practiced it since early childhood.

Capable humanoid robots are needed to investigate the above issues and for the deployment as a museum tour guide. This application requires interacting with multiple persons, most of them inexperienced in the interaction with robots. In this paper, we present the mobile humanoid museum tour guide robot Robotinho that we developed as successor to the communication robot Fritz [4]. Robotinho has a lightweight body with 25 major joints and is equipped with an expressive communication head. We detail mechanical and electrical design and give an overview of the perception of communication partners and of Robotinho's multimodal interaction capabilities. Furthermore, we describe the newly developed navigation modules that are needed for a scenario in which the robot acts as a tour guide.

This work has partly been supported by the German Research Foundation (DFG) under contract numbers BE 2556/2-2 and SFB/TR-8.

F. Faber, M. Bennewitz, C. Eppner, A. Görög, C. Gonsior, and D. Joho are with the Dept. of Computer Science, University of Freiburg, Germany.

M. Schreiber and S. Behnke are with the Dept. of Computer Science VI, University of Bonn, Germany.

Our robot uses speech, emotional expression, eye-gaze, and gestures to interact with people. Depending on the audio-visual input, our robot shifts its attention between different persons. Robotinho performs human-like arm gestures during the conversation and also uses pointing gestures generated with eyes, head, and arms to direct the attention of its interaction partners towards the explained exhibits (see Fig. 1). To express its emotional state, the robot generates facial expressions and adapts the speech synthesis. Robotinho is fully autonomous and it is equipped with a laser range finder and ultrasonic distance sensors for navigation in addition to two cameras and microphones for interaction. The robot is able to walk omnidirectionally while avoiding obstacles and to localize itself in a previously learned map of the environment.

The remainder of this paper is organized as follows. In the next section, we review the state of the art in human-humanoid interaction. In Sec. III, we present the mechanical and electrical design of Robotinho. Sec. IV covers the perception of communication partners and Sec. V details the multimodal interaction capabilities of our robot. Sec. VI describes the navigation system that we developed in order to enable the robot to give museum tours. Finally, we report experiences we made during public demonstrations of our system in Sec. VII.

Fig. 2. Our robot Robotinho was initially used as soccer player in the RoboCup Humanoid League. He conducted The 12 Cellists of the Berlin Philharmonic in September, 2006. The right image shows the communication head as well as the position of the cameras and distance sensors.

II. Related Work

Over the past century, robots became familiar figures in movies and TV series. The popularity of such robots indicates that people are receptive to the idea that these machines will one day become part of our everyday life.

Wheeled robots, for example, have already been deployed as museum tour guides or on large fairs [5], [6], [7], [8]. The main focus in these systems, however, was reliable, collision-free navigation. The researchers of these projects did not emphasize natural, human-like interaction.

In contrast to such systems, humanoid robots offer the potential to realize intuitive and more human-like interaction. Bischoff and Graefe [9] installed the robot Hermes for a long-term experiment in a museum. Hermes possesses an upper body with arms that is mounted on a wheeled base. However, the robot does not have an animated face and its multimodal interaction capabilities are limited. Shiomi *et al.* [10] deployed a set of four robots in a museum. Two of these were wheeled robots with a human-like upper body, the other two robots were small humanoid robots (construction kit). One of the wheeled robots served as a tour guide, the other robots interacted with the visitors or with themselves. The interaction was only limited, since speech recognition was neglected.

Kismet [11] is a robot head containing multiple cameras which has been developed for studying human-robot social interaction. It does not recognize the words spoken, but it analyzes low-level speech patterns to infer the affective intent of the human. Kismet displays its emotional state through various facial expressions, vocalizations, and movements. It can make eye contact and can direct its attention to salient objects. A more complex communication robot is Leonardo, developed by Breazeal *et al.* [1]. It has 65 degrees of freedom to animate the eyes, facial expressions, the ears, and to move its head and arms.

Leonardo and Kismet are mounted on a static platform. Mobile robots used for communication include PaPeRo [12], Qrio [13], Maggie [14], and Armar [15].

When designing robots for human-robot interaction, one must consider the uncanny valley effect, described by Mori [16]. Humans are no longer attracted to robots, if they appear too human-like. Photo-realistic android and gynoid robots, such as Repliee Q2 [2], are at first sight undistinguishable from real humans, but the illusion breaks down as soon as the robots start moving. For this reason, our robot does not have a photo-realistic human-like appearance, but we emphasize the facial features of its communication head using distinct colors.

III. Mechanical and Electrical Design

A. Mechanics

Our humanoid robot Robotinho, shown in Fig. 2, has been originally designed for playing soccer in the RoboCup Humanoid League TeenSize class. Robotinho is 110cm tall and has a total weight of about 6kg. Its body has 25 degrees of freedom (DOF): six per leg, four per arm, three in the trunk, and two in the neck. The mechanical design for the body focused on simplicity, robustness, and weight reduction. The body is driven by a total of 35 Robotis Dynamixel DX-117 intelligent actuators. All joints in the legs and the trunk, except for the yaw axes, are driven by two parallel actuators. They are used in a master-slave configuration to double the torque.

Robotinho's skeleton is constructed from aluminum extrusions with rectangular tube cross section. In order to reduce weight, we removed all material that was not necessary for stability. The feet and the forearms are made from sheets of carbon composite material. The elasticity of the feet helps to maintain non-degenerate foot-ground contact, even when the supporting foot is not parallel to the ground. The robot is protected by a layer of foam and an outer shell of thin carbon composite material.

For the use as communication robot, we equipped Robotinho with an expressive 15DOF head, visible in the right part of Fig. 2. All joints are driven by small digital servos. The eyes are movable USB cameras. Four servos move the eyes in two axis. While the lower eye lid moves together with the eyeballs, the upper eye lid can be moved independently in pitch direction. Six servo motors animate jaw and mouth. One pair of servos moves the jaw in pitch direction. Each mouth corner is moved by two servos and four servos animate the two eyebrows.

B. Electronics

Robotinho is fully autonomous. The robot is powered by high-current Lithium-polymer rechargeable batteries, which are located in its hip. Four Kokam 3200mAh cells last for about 25 minutes of operation.

The Dynamixel actuators have a RS-485 differential half-duplex interface. Robotinho is equipped with a CardS12 microcontroller board, which manages the detailed communication with all Dynamixels. The head servos are controlled using PWM, which is generated by two ChipS12 boards. In addition to these joint sensors, Robotinho is equipped with an attitude sensor, located in the trunk. It consists of a dual-axis accelerometer (ADXL203, $\pm$1.5g) and two gyroscopes (ADXRS 300, $\pm 300°$/s).

We use a tiny Sony Vaio UX PC as the main computer. The weight of the UX is only 0.5kg. It is attached to the lower back of Robotinho. The audio output of the UX is

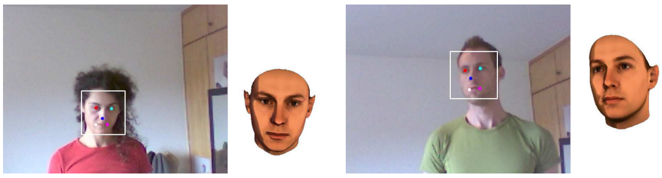

Fig. 3. Feature-based head pose estimation.

connected to a lightweight pair of active speakers, which are located in Robotinho's chest.

A mini-USB hub connects the two cameras (Videology 21K155) to the main computer. One camera has a narrow field of view of 21.4° while the other camera is equipped with a wide-angle lens, yielding a horizontal field of view of 64°. For localization and obstacle avoidance, Robotinho is equipped with a small laser range finder (LRF), located in its neck. The Hokuyo URG-04-LX has a weight of only 160g and produces 240° range scans at a rate of 10Hz. Its maximal measurement distance is 5.4m. The LRF is interfaced via the USB-hub to the main computer. Furthermore, we added a ring of eight ultrasonic distance sensors (Devantech SRF02) around the body at the height of the hip. The corresponding distance measurements are additionally used for the avoidance of the obstacles which are below the height of the laser.

IV. Perception of Communication Partners

A. Visual Perception

To detect and track people in the environment of our robot, we use the images of the two cameras. The different viewing angles of the cameras allow to cover a wider field of view and to detect people (i.e., their face) at larger distances, as compared to Robotinho's predecessor Fritz. Our robot maintains a probabilistic belief about the people in its surroundings which is updated based on detected faces in the camera images. Using this belief, the robot is also able to keep track of people when they are temporarily outside its field of view. The face detection and tracking system is described in detail in [17].

In order to recognize non-verbal signs of attention and intention we developed a module for head-pose estimation from monocular camera image sequences [18]. Starting from the detected faces, we search for distinctive facial features, such as the eyes, the nose tip, the mouth corners, and the ears. These features are localized using a boosted cascade of Haar-classifiers. We track these facial features and estimate the three Euler angles of head rotation around the neck using a neural function approximator. Input to the neural net are the relative feature positions within the bounding box of the face. Fig. 3 shows two examples with detected facial features and the corresponding estimated head pose.

Fig. 4. Gesture recognition.

Furthermore, we developed a system which is able to recognize typical human gestures such as head shaking/nodding, waving, or pointing given data of a monocular camera [19]. Also starting from the detected faces, we localize the hands by using an adaptive skin color model in combination with luminance-based hand detectors. We model the trajectories of the hands and the head using hidden Markov models. These HMMs are then used to recognize gestures online. For parametric gestures, such as pointing and size indicating gestures, we localize the hold phase (which carries the meaning) in time using the inferred HMM states. We then estimate the respective gesture parameter from the relative positions of the head and the hands during the hold phase. Fig. 4 shows two of the recognized gestures.

B. Auditory Perception

Speech is recognized using a commercial ASR system from Loquendo [20]. This system is speaker-independent and uses a small vocabulary grammar which is changed with the dialog state.

We also implemented a speaker localization system that uses a stereo microphone. We apply the cross-power spectrum phase analysis [21] to calculate the time difference between the left and the right channel. This yields an estimate of the horizontal angle between the speaker and the microphones.

V. Multimodal Interaction with Humans

A. Attentional System

Our robot shows interest in multiple persons in its vicinity and shifts its attention between them so that they feel involved into the conversation. To determine the focus of attention of the robot, we compute an importance value for each person in the belief, which is based on the time when the person has last spoken, on the distance of the person to the robot, and on its angular position relative to the front of the robot.

The robot always focuses its attention on the person who has the highest importance, which means that it keeps eye-contact with this person. While focusing on one person, from time to time our robot also looks into the direction of other people to involve them into a conversation and to update its belief.

Turning towards interaction partners is distributed over three levels [22]: the eyes, the neck, and the trunk. We use different time constants for these levels. While the eyes are allowed to move quickly, the neck moves slower, and the trunk follows with the slowest time constant. This reflects the different masses of the moved parts. When a saccade is made, the eyes point first towards the new target. As neck and trunk follow, the faster joints in this cascade move back towards their neutral position. A comfort measure, which incorporates the avoidance of joint limits, is used to distribute the twist angle over the three levels.

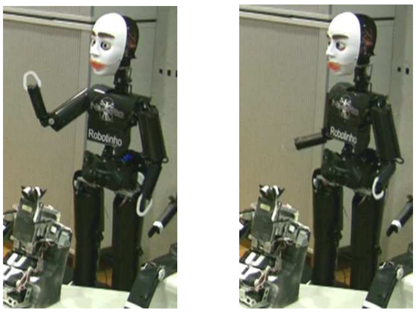

Fig. 5. While interacting with people, our humanoid robot performs several natural gestures.

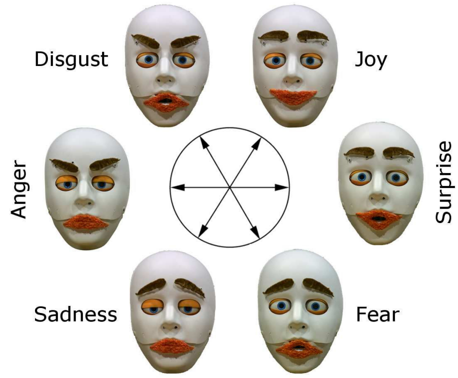

Fig. 6. Facial expressions used to display the emotional state of Robotinho.

B. Multimodal Dialog System

The dialog system covers a restricted domain only. The visitors can have some small talk with the robot and can tell it to give a museum tour. The level of detail of the explanation about exhibits can be influenced by the visitors and the robot is also able to answer some questions about itself. The dialog system is realized using a finite-state automaton. Some of the transitions are triggered by timeouts. The dialog system utilizes also the confidence reported by the speech recognition system and asks for confirmation in ambiguous cases.

C. Arm and Head Gestures

Our robot performs several natural, human-like gestures (see Fig. 5 and 1). These gestures either support its speech or correspond to unconscious arm movements which we humans also perform. The gestures are generated online. Arm gestures consist of a preparation phase where the arm moves slowly to a starting position, the hold phase that carries the linguistic meaning, and a retraction phase where the hand moves back to a resting position. The hold phase is synchronized with the speech synthesis module.

○ *Symbolic Gestures:* The symbolic gestures in our dialog system include a single-handed greeting gesture that is used while saying hello to newly detected people. The robot performs a come-closer gesture with both arms when detected persons are farther away than the normal conversation distance. There robot also accompanies certain questions with an inquiring gesture where it moves both elbows outwards to the back. In appropriate situations, the robot performs a disappointment gesture by moving, during the stroke, both hands quickly down. To confirm or to disagree, the robot also nods or shakes its head, respectively.

○ *Batonic Gestures:* Humans continuously gesticulate to emphasize their utterances while talking to each other. Robotinho also makes small emphasizing gestures with both arms when he is generating longer sentences.

○ *Pointing Gestures:* To draw the attention of communication partners towards objects of interest, our robot performs pointing gestures. When the robot wants to draw attention to an object, it simultaneously moves the head and the eyes in the corresponding direction and points in the direction with the respective arm while uttering the object name.

○ *Non-Gestural Movements:* Our robot performs small movements with its arms to appear livelier. We also implemented a regular breathing motion and pseudo-random eye blinks.

D. Emotional Expression

Showing emotions plays an important role in inter-human communication. During an interaction, the perception of the mood of the conversational partner helps to interpret his/her behavior and to infer intention.

• *Facial Expressions:* To communicate the robot's mood, we use a face with animated mouth and eyebrows to display facial expressions. The robot's mood is computed in a two-dimensional space, using six basic emotional expressions (joy, surprise, fear, sadness, anger, and disgust). Here, we follow the notion of the Emotion Disc developed by Ruttkay *et al.* [23]. This technique allows continuous changes of the facial expression. Fig. 6 shows how we implemented the six basic expressions for Robotinho using eyebrows, eyelids, mouth corners, and the jaw.

• *Emotional Speech Synthesis:* In combination with facial expressions, we use emotional speech to express the robot's mood. Most speech synthesis systems do not support emotional speech directly; neither does Loquendo TTS [20]. However, in this system, we can set the parameters average pitch, speed, and volume and thereby communicate the robot's emotional state.

VI. Navigation

A mobile robot needs to be able to walk to its goal pose without colliding with a static or dynamic obstacle. Furthermore, in order to know where to move to reach the next exhibit, the robot needs to learn a map of the environment first and has to be able to localize itself within this map. We use a combination of local and global navigation to achieve these tasks.

A. Collision-Free, Omnidirectional Walking

Robotinho is able to execute omnidirectional walking commands. The walking patterns are generated online [24]. Feedback from the gyroscopes is used to stabilize the trunk while walking [25]. The gait-target vector $\mathbf{v} = (v_x, v_y, v_\theta)$, which describes the desired walking speed in forward/backward, sidewards, and rotational direction can be set by the navigation module. The gait target can be changed smoothly while the robot is walking.

For local navigation to the next intermediate goal point, we use a potential field approach [26]. It continuously generates the gait target vector that guides the robot while avoiding obstacles sensed with the LRF and the ultrasonic distance sensors.

B. Self-Localization and Mapping

To enable autonomous navigation, Robotinho localizes itself via Monte Carlo localization [27] in grid maps given data of the LRF. We use the weighted mean of the particles representing the belief of the robot about its pose as the final pose estimate. To learn a grid map given laser range data collected with a walking humanoid, we apply the technique presented by Stachniss *et al.* [28]. This approach addresses the difficulties that appear especially in the context of humanoid robots. The key differences compared to wheeled robots are missing odometry information, comparably noisy data from light-weight proximity sensors, as well as a non-constant attitude (roll and pitch angle) resulting from the walking behavior.

C. Museum Tours

We use two classes of intermediate goal points for realizing museum tours. The first one are so-called waypoints which lie on the way between two exhibits and which are 2D positions. The idea is to "guide" the robot to the next exhibit. The second class of goal points are stop points which additionally have an orientation. A stop point is defined for each exhibit and this is the pose where the robot is supposed to stand while explaining the exhibit. We use an undirected graph of waypoints and apply the Dijkstra algorithm to find the shortest path to the next exhibit in this graph.

Note that we use a tolerance radius for the translation (0.5m) as well as for the orientation (5°) for reaching the stop points. Therefore, in this mobile scenario, we have to compute online the relative position of the exhibit to the robot since the robot does not always stop at exactly the same pose. We consider the estimated pose for generating the pointing gesture to draw the attention to the exhibit.

D. Wireless Communication

Because the payload of Robotinho is limited, we use an external computer for some of the less time-critical computations. The communication between the onboard PC and the external computer is implemented in the UDP protocol, which is transmitted over wireless LAN. The external computer runs speech recognition, self-localization, data logging, and visualization.

VII. Public Demonstrations

A. Conducting Cellists of the Berlin Philharmonic

Robotinho's first use as communication robot was in September 2006. At this time, it had a simple head with a static facial expression. As shown in the middle of Fig. 2,

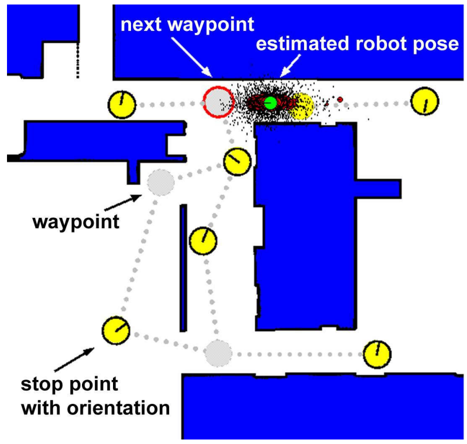

Fig. 7. Museum tour with stop points for exhibits and waypoints, which "guide" the robot through narrow passages and also serve as branching points. The particles representing the belief about the robot's pose and their weighted mean are also shown.

Robotinho conducted The 12 Cellists of the Berlin Philharmonic. This required extensive use of body language to express desired the playing speed, dynamics, cues, stops, etc. The robot also made full-body gestures such as a bow towards the audience and a standing-up request towards the cellists. A short video giving an impression about this demonstration can be found at our web page[1].

B. Explaining Exhibits

Our multimodal dialog system was tested during numerous lab demos. We chose a scenario in which the communication robot presents four of its robotic friends. We placed the exhibits on a table in front of the robot. Our communication robot interacted multimodally with the people and had simple conversations with them. Questionnaires filled out by the visitors indicate that most people found the eye-gazes, gestures, and the facial expressions human-like and felt that the robot was aware of them [4]. The visitors were also able to reliably dereference its pointing gestures. We provide a video showing Robotinho in the static scenario at our web page[2].

C. Guiding Visitors

The robot also gave tours in the corridor of our university building. The scenario is such that people can first have some small talk with the robot and then, as soon as someone shows interest in a tour, Robotinho guides the visitors to the exhibits and explains them. Fig. 1 shows Robotinho drawing attention to an exhibit after having guided the people to it. A video of Robotinho giving a tour can be found at our web page[3]. The people obviously enjoyed the interaction with Robotinho and followed the robot on the tour to the exhibits.

D. RoboCup@Home

Robotinho was part of the NimbRo team at the 2009 RoboCup@Home competitions where our team reached the

[1] www.nimbro.net/movies/robotinho/Robotinho_conducts_very_short.wmv
[2] www.nimbro.net/movies/robotinho/robotinho_static.wmv
[3] www.nimbro.net/movies/robotinho/robotinho_tourguide.wmv

Fig. 8. Robotinho guiding a guest through a home environment during the 2009 RoboCup@Home competition.

third place. In these competitions, robots operate in close cooperation with human users in a home environment. The competitions include a number of tests for different capabilities. Important aspects are navigation in a home environment and intuitive multimodal interaction with the users. Fig. 8 shows Robotinho giving a home tour to a guest during the 2009 RoboCup@Home competition. It should be noted that our team won also the innovation award for "Innovative robot body design, empathic behaviors, and robot-robot cooperation".

VIII. Conclusions

In this paper, we presented the mobile full-body humanoid museum tour guide Robotinho. In contrast to its predecessor Fritz, Robotinho is not only used in a static scenario and it possesses several new features such as a movable trunk, a more expressive head with movable eye lids, and an additional arm joint.

Similar to Fritz, Robotinho uses multiple modalities to interact with people in an intuitive, natural way. These include speech, emotional expressions, eye-gaze, and a set of human-like, symbolic as well as unconscious arm and head gestures.

To go from a static exhibit explanation scenario to a mobile museum tour guide scenario, we implemented omnidirectional walking, self-localization, mapping, obstacle avoidance, and path planning for Robotinho. As we reported in the experimental section, the robot is able to give a tour at our university building and to interact with untrained users in a natural way. These skills were also well received by the jury in the RoboCup 2009 @home competitions.

References

[1] C. Breazeal, A. Brooks, J. Gray, G. Homan, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, and D. Chilongo, "Tutelage and collaboration for humanoid robots," *Int. J. of Humanoid Robotics*, 2004, 1(2):315-348.

[2] D. Matsui, T. Minato, K. F. MacDorman, and H. Ishiguro, "Generating natural motion in an android by mapping human motion," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2005.

[3] T. Spexard, A. Haasch, J. Fritsch, and G. Sagerer, "Human-like person tracking with an anthropomorphic robot," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), Orlando, FL*, 2006, pp. 1286–1292.

[4] M. Bennewitz, F. Faber, D. Joho, and S. Behnke, "Fritz - A humanoid communication robot," in *Proc. 16th IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, 2007, pp. 1072–1077.

[5] I. Nourbakhsh, C. Kunz, and T. Willeke, "The Mobot museum robot installations: A five year experiment," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2003.

[6] R. Siegwart, K. Arras, S. Bouabdallah, D. Burnier, G. Froidevaux, X. Greppin, B. Jensen, A. Lorotte, L. Mayor, M. Meisser, R. Philippsen, R. Piguet, G. Ramel, G. Terrien, and N. Tomatis, "Robox at Expo.02: A large-scale installation of personal robots," *Robotics & Autonomous Systems*, vol. 42, no. 3-4, pp. 203–222, 2003.

[7] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, J. Schulte, and D. Schulz, "Probabilistic algorithms and the interactive museum tour-guide robot Minerva," *Int. J. of Robotics Research (IJRR) 19(11):972-999*, 2000.

[8] W. Burgard, A. 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.

[9] R. Bischoff and V. Graefe, "Demonstrating the humanoid robot HERMES at an exhibition: A long-term dependability test," in *Proc. of the IROS Workshop on Robots at Exhibitions*, 2002.

[10] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita, "Interactive humanoid robots for a science museum," *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 25–32, 2007.

[11] C. Breazeal, *Designing Sociable Robots*. Cambridge, MA: MIT Press, 2002.

[12] O. Shin-Ichi, A. Tomohito, and I. Tooru, "The introduction of the personal robot papero," *IPSJ SIG Notes*, no. 68, pp. 37–42, 2001.

[13] K. Aoyama and H. Shimomura, "Real world speech interaction with a humanoid robot on a layered robot behavior control architecture," in *Proc. of ICRA*, 2005, pp. 3814–3819.

[14] J. Gorostiza, R. B. A. Khamis, M. Malfaz, R. Pacheco, R. Rivas, A. Corrales, E. Delgado, and M. Salichs, "Multimodal human-robot interaction framework for a personal robot," in *Proc. 15th Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, 2006.

[15] R. Stiefelhagen, H. K. Ekenel, C. Fügen, P. Gieselmann, H. Holzapfel, F. Kraft, K. Nickel, M. Voit, and A. Waibel, "Enabling multimodal human-robot interaction for the karlsruhe humanoid robot," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 840–851, 2007.

[16] M. Mori, "Bukimi no tani [the uncanny valley]," *Energy*, vol. 7, no. 4, p. 3335, 1970.

[17] M. Bennewitz, F. Faber, D. Joho, S. Schreiber, and S. Behnke, "Integrating vision and speech for conversations with multiple persons," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2005.

[18] T. Vatahska, M. Bennewitz, and S. Behnke, "Feature-based head pose estimation from images," in *Proc. of the IEEE-RAS Int. Conference on Humanoid Robots (Humanoids)*, Pittsburgh, USA, 2007.

[19] T. Axenbeck, M. Bennewitz, S. Behnke, and W. Burgard, "Recognizing complex, parameterized gestures from monocular image sequences," in *Proc. of the IEEE/RSJ Int. Conf. on Humanoid Robots (Humanoids)*, 2008.

[20] Loquendo S.p.A., "Vocal technology and services," http://www.loquendo.com, 2007.

[21] D. Giuliani, M. Omologo, and P. Svaizer, "Talker localization and speech recognition using a microphone array and a cross-powerspectrum phase analysis," in *Proc. ICSLP*, 1994, pp. 1243–1246.

[22] F. Faber, M. Bennewitz, and S. Behnke, "Controlling the gaze direction of a humanoid robot with redundant joints," in *Proc. 16th Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, 2008.

[23] Z. Ruttkay, H. Noot, and P. ten Hagen, "Emotion Disc and Emotion Squares: Tools to explore the facial expression space," *Computer Graphics Forum*, vol. 22, no. 1, pp. 49–53, 2003.

[24] S. Behnke, "Online trajectory generation for omnidirectional biped walking," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2006, pp. 1597–1603.

[25] F. Faber and S. Behnke, "Stochastic optimization of bipedal walking using gyro feedback and phase resetting," in *Proc. of the IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, Pittsburgh, USA, 2007.

[26] J. Barraquand, B. Langois, and J. Latombe, "Numerical potential field techniques for robot path planning," *IEEE Tr. on Robotics and Automation, Man and Cybernetics*, vol. 22, no. 2, pp. 224–241, 1992.

[27] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte carlo localization for mobile robots," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 1999.

[28] C. Stachniss, M. Bennewitz, G. Grisetti, S. Behnke, and W. Burgard, "How to learn accurate grid maps with a humanoid," in *Proc. of the IEEE Int. Conference on Robotics & Automation (ICRA)*, 2008.