# Robots meet humans

## Interaction in public spaces

# Robots Meet Humans—Interaction in Public Spaces

Björn Jensen, *Member, IEEE*, Nicola Tomatis, Laetitia Mayor, Andrzej Drygajlo, *Member, IEEE*, and Roland Siegwart, *Senior Member, IEEE*

***Abstract*—This paper presents experiences from *Robotics*, a long-term project at the Swiss National Exposition *Expo.02*, where mobile robots served as tour guides. It includes a description of the design and implementation of the robot and addresses reliability and safety aspects, which are important when operating robots in public spaces. It also presents an assessment of human–robot interaction during the exhibition. In order to understand the objectives of interaction, the exhibition itself is described. This includes details of how the human–robot interaction capabilities of the robots have evolved over a 5-month period. Requirements for the robotic system are explained, and it is shown how the design goals of reliability and safe operability, and effective interaction, were achieved through appropriate choice of hardware and software, and the inclusion of redundant features. The modalities of the robot system with interactive functions are presented in detail. Perceptive elements (motion detection, face tracking, speech recognition, buttons) are distinguished from expressive ones (robotic face, speech synthesis, colored button lights). An approach for combining stage-play and reactive scenarios is presented. The authors also explain how an emotional state machine was used to create convincing robot expressions. Experimental results, both technical and those based on a visitor survey, as well as a qualitative discussion, give a detailed report on the authors' experiences in this project.**

***Index Terms*—Human–robot interaction, mobile robot, modalities for interaction, public space experience.**

## I. Introduction

MOBILE robots have begun to appear in public spaces such as supermarkets, museums, and expositions. These robots need to interact with people and to provide them with information. They have to invite people to use the services offered. To do so, communication must be intuitive, so that people, inexperienced with mobile robots, can interact with the system without prior instructions. This calls for spoken dialogues, as it is the natural means of communication among us.

Tour-guide robots are required to perform in dynamic environments. This often involves responding to complex inputs from several sources. In other words, sensory interpretation and action preparation become primal aspects of such systems. Their action–perception loop should detect and register several kinds of events and create appropriate motion and expressions.

Manuscript received February 17, 2004; revised August 19, 2004. Abstract published on the Internet September 26, 2005. This work was supported by *Expo.02* and EPFL.

B. Jensen, A. Drygajlo, and R. Siegwart are with the Swiss Federal Institute of Technology, Lausanne CH-1015, Switzerland (e-mail: bjoern.jensen@epfl.ch; andrzej.drygajlo@epfl.ch; roland.siegwart@epfl.ch).

N. Tomatis is with the Swiss Federal Institute of Technology, Lausanne CH-1015, Switzerland and also with BlueBotics SA, Lausanne CH-1015, Switzerland (e-mail: nicola.tomatis@epfl.ch).

AQ1 L. Mayor is with Helbling Technik, AG.

Digital Object Identifier 10.1109/TIE.2005.858730

At the Swiss National Exhibition *Expo.02*, 11 RoboXs were used as tour guides in a public exposition for a period of five months. Presentation and reactive scenarios are combined using stage-play elements and a continuously running emotional state machine. Reactive scenarios were used in the events of obstruction, wrong use of interaction modalities by the user, and low battery level.

Tour guiding required the robots to move in a densely populated exposition space from exhibit to exhibit. Closeness to the visitors called for safe operation of the robot. The long duration of the exposition made system reliability an important design goal. Requirements for human intervention and supervision had to be kept within tight limits, in order to make the *Robotics@Expo.02* a success, and to render interaction credible.

### A. Structure

This paper has three goals, namely: 1) describing design and construction elements required to achieve reliable and safe operation during the *Expo.02*; 2) presenting modalities and strategies for interaction; and 3) assessing the interactive performance achieved by the tour-guide robot.

After reporting on related work, the exposition *Expo.02* is outlined. The tour-guide robot is presented and its modalities for interaction are explained. The creation of interactive scenarios is addressed and the functioning of the emotional state machine is explained.

Results comprise the performance of the robot and of its individual modalities for interaction and a survey on human–robot interaction. To conclude, experiences from operating the robots during the 5-month period are summarized as a qualitative discussion of the evolution of interaction scenarios.

### B. Related Work

There are a variety of robotic systems for interaction, some of which are commercialized (e.g., Sony's AIBO [1]) or at a prototype stage (e.g., Honda's ASIMO [2]), while others are used in research and academia. They underline the importance of appearance, which has to be sufficiently lifelike, while still remaining distinctly artificial. In order to avoid the uncanny valley [3] of emotional rejection, such systems should be well received by the user. This is emphasized as well by Kismet [4], a robot research platform able to learn behavior. In these cases, interaction is a reactive task, usually involving one human and one robot.

Among the publications pertaining to robots in expositions, some focus on navigation [5]–[7], while others stress on the interaction modalities [8]–[10].

By navigational aspects, we mean the task of guiding visitors, particularly in densely populated environments: maintaining visitor interest and allowing a group to move toward the next exhibit by asking for leeway in situations where the robot is blocked.

Experience with Rhino [5] in public spaces underlines the importance of dedicated interfaces for interaction. The tour-guide robot Minerva [6] was equipped with a face and had four different emotional states to further improve interaction. The navigation approach of these robots has shown its strength in museums for one week (19 km) and two weeks (44 km), respectively. This navigation relied on off-board resources and is reported to be sensitive to environmental dynamics.

The Mobot Museum Robot Series reported in [8] and [9] puts more focus on interaction and design, simplifying its navigation task by means of artificial landmarks in the environment. The robots Sage [8], Chips, Sweetlips, Joe, and Adam [9] emerged over the years and used an increasing number of interaction modalities. They operated for up to three years with Sage covering a total of 323 km [8] and Chips, Sweetlips, Joe, and Adam each covering more than 600 km [9]. With the exception of the last mentioned robot, the movements of the others were limited to a predefined set of unidirectional safe routes in order to simplify both localization and path planning.

More expressive modalities do not necessarily imply better interaction. In [11] and [12], the effectiveness of several modalities for interaction is evaluated based on the attention that a robot receives. The human interest in a robot also varies over time, as a school class experiment [13] shows. In the beginning, the unusual robot experience raised enormous interest among the pupils that vanished within a week. Apparently, success in short term and long term have different reasons and may require different modalities.

Another permanent installation of mobile robots is at the Deutsches Museum für Kommunikation (German Museum of Communication) in Berlin [7]. Three robots have a dedicated task each, like welcoming visitors, offering them exhibition-related information, and entertaining visitors. They navigate in a restricted and structured area. Localization uses segment features and a heuristic scheme for matching and pose estimation. Information about the museum is provided using multimedia equipment, and one robot chases a ball.

### C. *Expo.02*

The Swiss National Exhibition takes place approximately once every 40 years. *Expo.02* took place from May 15 to October 21, 2002. It was a major national happening with 37 exhibitions and an event-rich program. The *Robotics@ Expo.02* exhibition [14] was intended to show the increasing closeness between humans and robots. The central visitor experience of *Robotics@Expo.02* was the interaction with autonomous freely navigating mobile robots giving guided tours and presenting the exhibits shown in Fig. 1. The exhibition was scheduled for a visitor flow of 500 persons per hour. The average duration of a complete tour of the 315 $m^2$ exposition area was planned for 15 min.

After agreeing on one of the official languages of *Expo.02* (English, French, Italian, or German), the robot started moving to the exhibits like Industry robot (A), Medical robot (B), Fossil (D) (showing body implants), or mechanical underwater toys at Aquaroids (E). Visitors could control the miniature robot Alice (F) using buttons on the tour-guide robots. Other exhibits like Face Tracking (K) and our Supervision Lab (M) or the robot presentation of itself Me, myself and I (C) gave some insight to the mobile robots' perception of the environment.

The tours were dynamic, in that the exhibits presented were chosen by the visitor. After completing the presentation of one exhibit, robots requested a list of free exhibits. To promote visitor flow toward the exit, only free exhibits, located closer to the exit than the current could be selected by the visitors. A tour ended after a fixed number of exhibits, with the robot saying goodbye and returning to the welcome area.

Some robots were dedicated to one exhibit and interacted without the need to give a tour: the Presenter robot (G), explaining the inner workings of a robot, the Jukebot (H), proposing a selection of music, the Philosopher (J), speaking about good and the world, and the Photographer (L), taking pictures and displaying them on three television towers, the so-called Cadavre Exquis (N).

## II. Tour Guide: RoboX

The autonomous mobile system RoboX was developed for *Expo.02* at the Autonomous Systems Lab and produced by its spin-off company BlueBotics SA. It is shown in Fig. 2. Safe and reliable operation was mandatory for its use in a public exposition, in close proximity to hundreds of visitors. For most of the visitors, RoboX was the first contact with a real robot. This called for friendly appearance and an intuitive operation. How visitors would react toward an autonomous machine was difficult to predict. Thus, considerable effort was undertaken to make the robot robust against destructive behavior.

### A. *Hardware*

In order to ensure that visitors could easily spot RoboX even in crowded settings, the robot's height is 1.65 m. Heavy components are in its mobile base, which has a diameter of 0.70 m (0.90 m with foam bumpers), giving the robot good equilibrium. The battery pack provides up to 12 h of autonomy and makes up a large part of the system's weight of 115 kg. RoboX has two differentially driven wheels on its middle axis, which allows turning on the spot. This is a key feature when visitors are blocking its way.

The mobile base contains the following: two laser range finders (Sick LMS 200); the drive motors; the safety circuit; and the tactile bumpers. Additionally, the two computers making the robot autonomous, a PowerPC 375 MHz running XO/2 and a personal computer (PC) Pentium III 700 MHz running Windows 2000, are located there. To interact with visitors, RoboX provides a mechanical face with a Firewire color camera and a light-emitting diode (LED) matrix, two loudspeakers, and interactive buttons. Two robots were equipped with a directional microphone matrix (Andrea Electronics DA-400 2.0) for

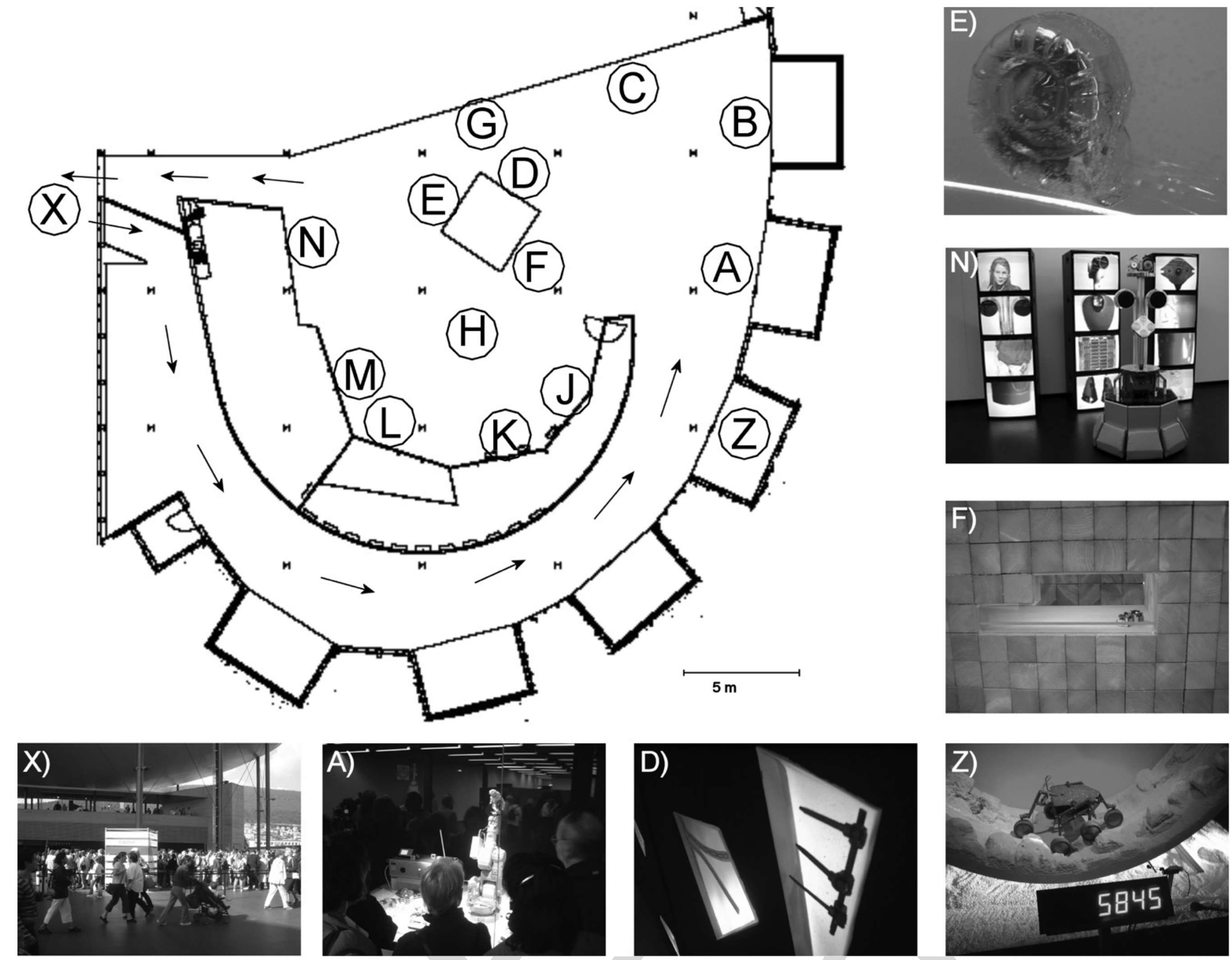


Fig. 1. Overview of the *Robotics* exhibition at *Expo.02*. The plan in the upper left indicates the location of exhibits and other places of interest. The insets are labeled accordingly, as well as some references in the main text. Exhibits A–N were parts of guided tours (exhibit Z was added to this list for the last two months). Label X denotes the exit. (A) Industrial robot playing with toys. (B) Medical robot. (C) Me, myself and I. (D) Fossil (medical implants in amber). (E) Aquaroids (underwater toys). (F) Alice, the sugar-cube sized minirobot. (G) Presenter robot. (H) Jukebot. (J) Philosopher. (K) Face Tracking. (L) Photographer. (M) Supervision Lab. (N) Cadavre Exquis mixing photos of visitors taken by Photographer with images of mechanical parts in order to create virtual cyborgs. (X) Exposition seen from the outside. (Z) Shrimp, the outdoor robot in a huge hamster wheel.

speech recognition. Modalities for interaction are explained in more detail in Section III.

### B. Navigation

The navigation system is composed of localization, path planning, and obstacle avoidance. These tasks are executed by the real-time operating system (RTOS) running on the PowerPC. No off-line resources are required. A graph-based *a priori* map underlies localization and global path planning. It contains geometric and topological information. Exhibits are represented as goal nodes. Via nodes, which are nodes with a bigger goal area, are used to model environment topology and anchor geometric features. A local geometric environment model is used for local path planning and obstacle avoidance.

Localization is based on line features extracted from laser range data, with multiple hypotheses tracked using a Kalman filter [15]. It was designed for operation in unmodified environments and performs well in cluttered situations. Using line features keeps the map compact and computational costs low.

Motion control combines several approaches, in a manner similar to the following [16]: NF1 [17] for local path planning; elastic bands [18] as adaptive path representation; and the dynamic window approach [19] for obstacle avoidance. The method has high computational efficiency due to lookup tables similar to [20]. More details can be found in [21].

### C. Safety

Robot components that influence motion are defined as safety critical, namely: speed control; obstacle avoidance; laser scanner; and bumpers. All those are running on the RTOS of the PowerPC. Taking into account the possibility of a failure of the PowerPC, a redundant safety controller is added. It is implemented using a peripheral interface controller

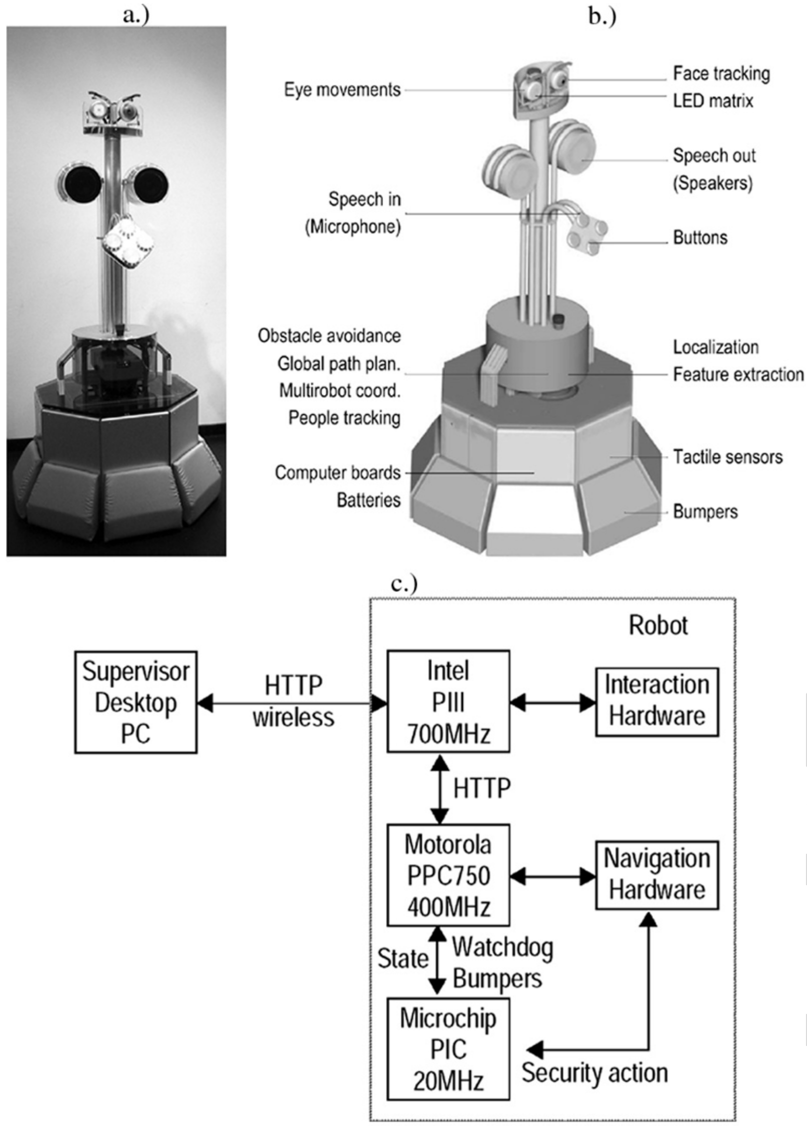


Fig. 2. (a) Interactive mobile robot RoboX. (b) Navigation and interaction elements of RoboX. (c) RoboX safety system layout: Navigational components on the RTOS of the PowerPC, Windows 2000 contains interactive components only (i.e., not safety critical). The PIC microcontroller serves as a watchdog and provides redundancy, it causes emergency stops in case of failures. Centralized supervision eases management of the 11 robots.

(PIC) microcontroller. In addition, centralized monitoring helps managing the 11 robots. The resulting system layout is shown in Fig. 2. RoboX also features a prominent emergency button to allow human intervention at all times.

Safety critical software runs under XO/2 on the PowerPC, a deadline-driven hard RTOS [22] designed for safe operation. Failure to execute a process within the required deadline causes the system to stop in a controlled manner.

In order to ensure safety in the event of failures in XO/2, the PowerPC, or related hardware, the PIC serves as a watchdog for several components. Speed control, obstacle avoidance, and laser scanner driver all emit watchdog signals verified by the PIC. Bumper contact requires an acknowledge signal from the PowerPC within only a small delay. If any of these signals is not received, or if the wheel speed exceeds 0.6 m/s, the PIC stops all robot motions by shorting the phases of the main actuators and sounds the alarm (light and sound).

## III. Modalities for Interaction

In an exhibition, the tour-guide robot interacts with individual visitors as well as crowds of people. In both situations, it is important that RoboX takes the initiative. Thus, a primary component of a successful tour guide is the ability to engage in a meaningful conversation in an appealing way [23]. High-performance environmental perception and intuitive expressive elements are the means used to achieve this goal.

In the following, the modalities for interaction are presented and their main features described. We distinguish perceptive and expressive modalities.

### A. *Perceptive Modalities*

RoboX is equipped with multiple sensors. A camera and two laser scanners give the robot a sense of people surrounding it, an important skill for interaction as reported in other public space

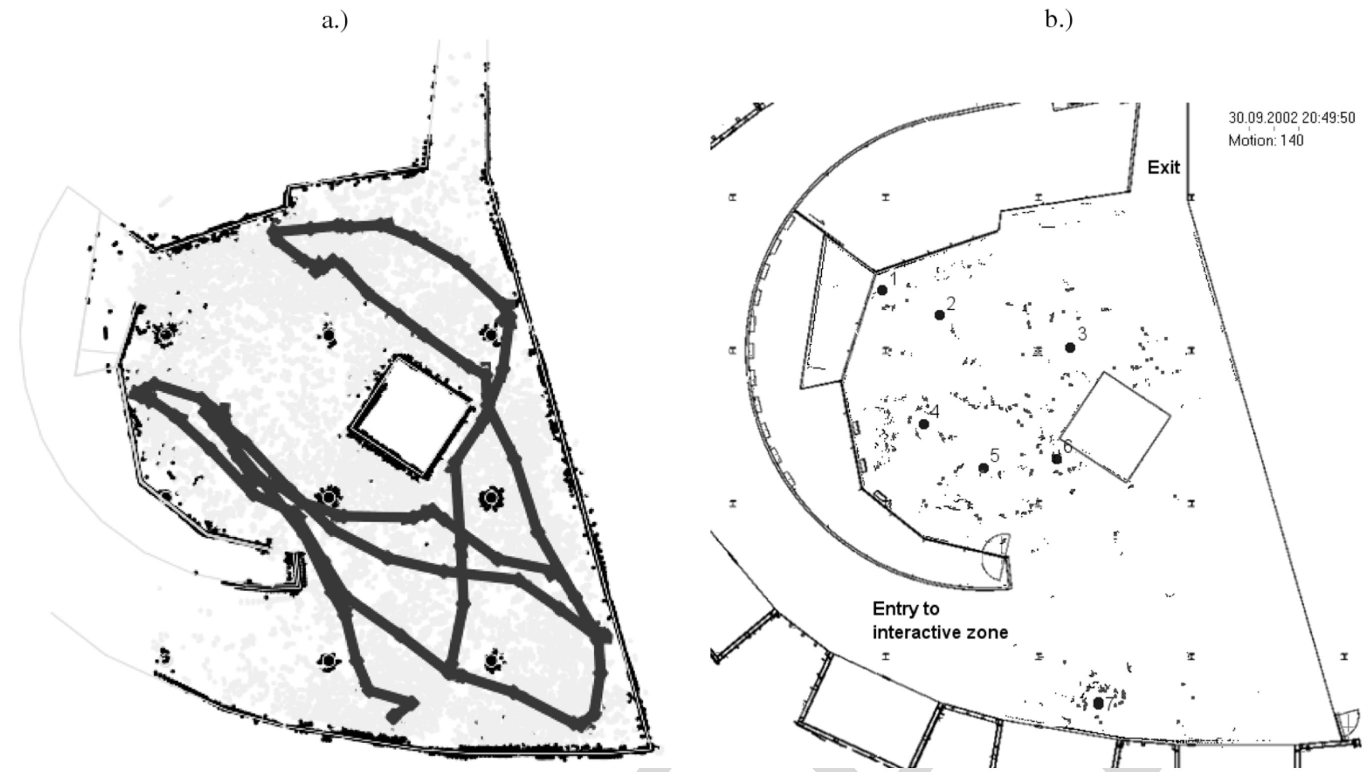


Fig. 3. Motion detection using laser range finder data from a mobile platform at *Expo.02* while roaming the 315 $m^2$ exhibition area. (a) The path of the robot during 17 min with light points indicating dynamic parts and dark points representing static parts. (b) Snapshot of the exposition with data from several robots. One hundred forty motion elements are detected at this moment.

experiences [5], [8], [9], [24]. The face tracking system detects the number of faces in the camera's field of view and determines how long they remain in front of the robot. Visitors use speech recognition or the buttons to interact with the robot. The robot also detects if someone or something touches the buttons or bumpers. Finally, the battery level is measured and used as an input for reactive scenarios and the emotional state machine.

In the following, the main perceptive elements are described in more detail.

*1) Motion Detection:* Motion is detected in order to find people in the robot's vicinity. Other methods could be employed, e.g., using shape information [25], [26] or singularities in the environment [27]. Our method is presented in detail in [28]–[30].

A result of the algorithm is shown in Fig. 3(a). The environment is assumed to be convex and static in the beginning. The range readings are integrated into the so-called static map, consisting of all currently visible elements that do not move. Only one information is stored for each angle. In the next step, the new information from the range finder is compared with the static map. Assuming a Gaussian distribution of the sensor readings representing a given element, a chi-square test can be used to decide whether the current reading belongs to one of the elements of the static map or originates from a dynamic object. All static readings are used to update the static map. Readings labeled as dynamic are used to verify the static map as follows: If the reading labeled as dynamic is closer to the robot than the corresponding value from the static map, the latter persists. In case it is farther away than the map value, it is used to update the map, but remains labeled as dynamic. All dynamic elements are clustered according to their spatial location. Each cluster is assigned a unique identification (ID) and the center of gravity of its constituting points in Cartesian space is computed. The classification, update, and validation steps are repeated for every new scan. In case of robot motion, the static map is warped to the new position.

*2) Face Tracking:* Fig. 4 shows an example of face tracking based on red green blue (RGB) data of the camera located in the robot's left eye. Skin-colored regions are extracted using an algorithm presented in [31] and [32]. To reduce the sensitivity against illumination, green and blue are normalized using the red channel. Then, fixed ranges for blue, green, and brightness are accepted as skin color. Taking brightness into account rejects regions of insufficient saturation. Erosion and dilation remove small regions from the resulting binary image. The binary image is clustered and the contour of each cluster is extracted. Heuristic filters are applied to suppress skin color regions that are not faces. These filters are based on rectangular areas, their aspect ratio, and the percentage of skin color within the rectangle. Clusters are linked over time using the nearest-neighbor assignment. Clusters that remain unassigned to previous tracks are added and tracked until they leave the camera's field of view.

Information gathered from the face tracker is used in several interaction parts. Together with motion tracking, it helps to verify the presence of visitors and to orient the robot's face toward the user. Furthermore, it triggers the behavior engine emotional state machine, which is presented in Section IV.

Fig. 4. Sequence of faces tracked by a RoboX at the *Robotics* exposition. From left to right and top to bottom, RoboX first tracks the face of a woman, then in the third image, it moves the eyes toward a man and tracks him until the next eye movement in the third image of the second row, where a third person appears.

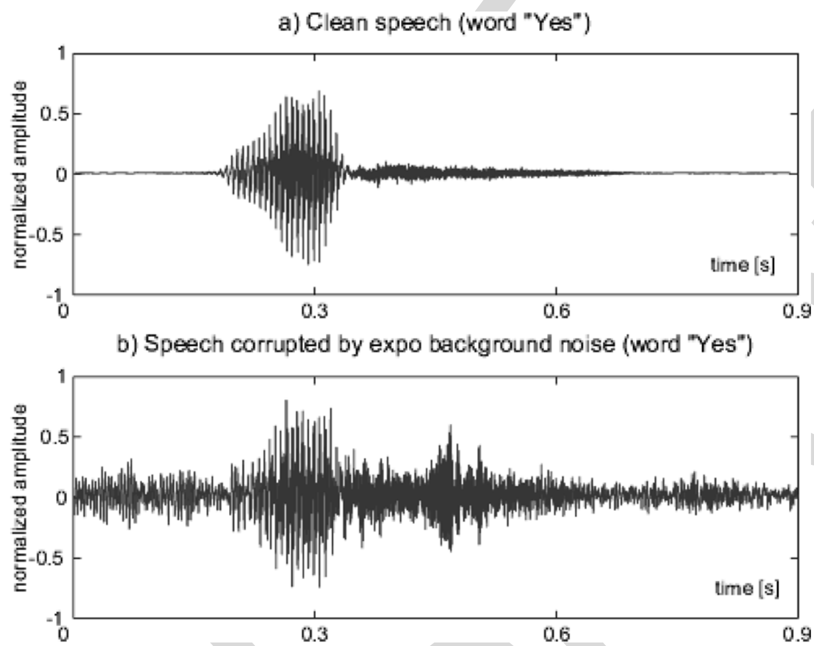


Fig. 5. Samples of the word Yes under (a) quiet and (b) noisy conditions of the exhibition room.

*3) Speech Recognition:* A primary requirement of *Expo.02* was that the tour-guide robots should be capable to interact with visitors using four languages: French, German, Italian, and English. The large number of visitors prohibited the use of handheld microphones as in [10], the adopted solution was to mount a microphone array on the robot.

Studying related work on tour-guide robots led us to the following observations [33]. First, even without voice-enabled interfaces, tour-guide robots are very complex, involving several subsystems that need to communicate efficiently in real time. This calls for speech interaction techniques that are easy to specify and maintain, and that lead to robust and fast speech processing. Second, the tasks that most tour-guide robots are expected to perform typically require only a limited amount of information from the visitors [34]. These points argue in favor of a very limited but meaningful speech recognition vocabulary and for a simple dialogue management approach. The solution adopted is based on yes/no questions initiated by the robot where visitors' responses can be in the four required languages (oui/non, ja/nein, si/no, yes/no). This simplifies the voice-enabled interface by eliminating the specific speech understanding module and allows only eight words as multilingual universal commands. The meaning of these commands depends on the context of the questions asked by the robot. A third observation is that tour-guide robots have to operate in very noisy environments, where they need to interact with many casual persons (visitors). Fig. 5 presents typical speech samples from quiet and noisy conditions. In the exhibition room, the signal is drowned in babble combined with the noise of robot movement and beep sounds. This calls for speaker-independent speech recognition and for robustness against noise. The first task of the speech recognition event is the acquisition of the useful part of the speech signal. The adoption of acquisition limited in time (3 s) is motivated by the average length of yes/no answers. Ambient noise in the exhibition room is among the main reasons for speech recognition performance degradation. A microphone array (Andrea Electronics DA-400 2.0) is used to add robustness without additional computational overhead. During the 3-s acquisition time, the original acoustic signal is processed by the microphone array. The mobility of the tour-guide robot is very useful for this task since the robot, when using the motion detection system, can position its front in the direction of the closest visitor and, thus, directs the microphone array. The preprocessing of signals of the array includes spatial filtering, dereverberation, and noise canceling. This preprocessing does not eliminate all the noise and out-of-vocabulary (other than yes/no) words. It provides sufficient quality and nonexcessive quantity of data for further processing. Recognition should perform equally well on native and foreign speakers of the target language. We are interested in a low error rate and rejection of irrelevant words. At the heart of the robot's speech recognition system lies a set of algorithms for training statistical models of words subsequently used for the recognition task. The signal from the microphone array is processed using a Continuous Density Hidden Markov Model (CDHMM) technique where feature extraction and recognition using the Viterbi algorithm are adapted to a real-time execution. It offers the potential to build word models for any speaker using one of the mentioned languages and for any vocabulary from a single set of trained phonetic subword units. The major problem of a phonetic-based approach is the need for a large database required for training a set of speaker-independent and vocabulary-independent phoneme models. This problem was solved using standard European and American databases available from our speech processing laboratory, as well as specific databases with the eight keywords recorded during experiments. Four language-specific databases were used to train four sets of phoneme-based subword models. Training employed the CDHMM toolkit HTK [35] based on the Baum–Welch algorithm. Out-of-vocabulary words and spontaneous speech phenomena like breath, coughs, and all other sounds that could cause a wrong interpretation of visitor's input also have to be detected and excluded. For this reason, a word spotting algorithm with garbage models has been added to the recognition system. These garbage models were built from the same set of phoneme-based subword models [36], [37], thus, avoiding an additional training phase or software modification. Finally, the basic version of the system was capable of recognizing yes/no words in the required languages and acoustic segments (undefined speech input) associated with the garbage models.

*4) Buttons:* Buttons were used as a robust means of enabling communication with the visitors under exposition conditions. They allow selecting the language, responding to questions, controlling exhibits via RoboX, and other types of actions. Their state (waiting for input, yes/no, language selection, etc.) was indicated by lights, making it an expressive component as well as an input device.

### *B. Expressive Modalities*

When RoboX finds people in close distance, it should greet and inform them of its intentions and goals. The most natural and appealing way to do this is by speaking. In addition to speech, a large number of facial expressions and body movements are used in human communication to enhance the meaning of the spoken dialogue. Additional expression is conveyed by varying prosodic parameters.

Fig. 6. Face mimicking the expressions joy, surprise, and disgust.

Certain researchers state that in order to socially interact with humans, robots must be believable and lifelike, must have behavioral consistency, and have ways of expressing their internal states [38]. Our goal was to create a credible character in that sense for guiding tours. We describe how the robot uses its face and speech synthesis to convey expressions.

*1) Face:* Communicating with humans usually seek the face of the dialogue partner. Its expressions provides crucial additional information for interpreting the spoken messages. To provide a similar anchor of communication for RoboX, the mechanical face, shown in Fig. 6 was built with two eyes. Expressions are created with its five degrees of freedom and the LED matrix in the right eye. Each eye has two degrees of freedom. The eyebrows have one common degree of freedom. There is no articulated mouth, to avoid synchronization problems with synthesized speech or the strange situation of a robot that speaks without moving its mouth.

The LED matrix displays small icons or animations. The matrix consists of 69 blue LEDs and serves as a miniature screen. It improves otherwise less comprehensible expressions. An intuitive way of conveying the robot's mood is changing the light intensity: Low light intensity makes the robot seem sad or tired, whereas bright light emits an impression of alertness. Expressiveness was achieved with eye movements and LEDs in two manners, namely: 1) showing an iris; or 2) displaying icons. The default picture on the matrix is the iris, its size is determined by the robot's mood. This creates a symmetric face since the left eye with the camera has a blue iris, too. The nondefault pictures are six icons that symbolize the six basic expressions (see Section IV), some of which are shown in Fig. 6. They appear at the same time as random eye movements intended to avoiding an uncomfortable robotic stare.

The LED display and eye movements express the state of the robot. Apparition effect, duration, and disappearance effect can be individually defined for each icon. Default expressions can be used for stage-play scenarios, i.e., when the robot executes a predefined sequence of movements to convey its internal state (Fig. 7). AQ2

*2) Speech Synthesis:* Speech synthesis allows the robot to express itself in the four languages of *Expo.02*. Environmental conditions (large rooms with many people) were a challenge for audibility.

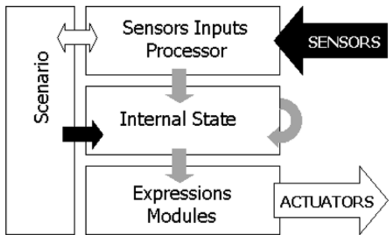


Fig. 7. Information flow: The scenario program is executed and influenced by sensor input. The internal emotional state is influenced by signals from several sources, including the scenario. RoboX expression results as a function of its internal state.

The use of prerecorded samples was ruled out by the requirement of conveying the robot's emotional state by modulating speech parameters, and to allow dynamic generation of spoken sequences. RoboX employs speech synthesis system based on LAIPTTS [39], [40] and Mbrola [41] for French and German, whereas English and Italian were synthesized using ViaVoice [42]. Prosodic parameters as pitch, volume, and rate can be changed while the robot is speaking.

## IV. Emotional State Machine

The emotional state machine is an internal representation modeling the mood of RoboX [43]. Its inputs are signals from several sources, including commands from the scenario. These change the internal emotional state, which is then mapped onto parameters of the modalities controlling the expression. It is not feasible to define all possible nuances explicitly. Therefore, we use a set of template expressions and derive displayed expressions through interpolation.

In the following, we describe how a set of template expressions is created; how signals from several sources influence the emotional state; how the emotional state is represented; and how this state is mapped on the modalities to create expressions.

### *A. Template Expressions*

Six template expressions are defined for the following: sadness; disgust; joy; anger; surprise; and fear. In addition, we define a neutral expression a calm state. The calm state proved particularly helpful for transitions from one expression to another.

For each template expression, a parameter set for the expressive modalities was defined manually. Table I shows the parameter sets qualitatively. We chose to mimic human expressions and to exaggerate them where possible, given the capacities of the robot.

To create a more lively appearance, these template expressions allow the definition of a value range for the expressive parameters. Within this range, the actual output is defined randomly and changes continuously. The emotional state machine provides the scenario with a control on how these parameter ranges are used:

1) Default behavior: Only eyebrows are controlled by the emotional state machine. Their position is changed according to the robot's current state.

TABLE I

Parameter Sets of Expressive Modalities for Template Expressions, With Small (S), Medium (M), Large (L), and Slow or Fast. Symbols (-?-) and (-X-) are Shown on the LED Matrix

| | Eye | | Speech | | |
|---|---|---|---|---|---|
| | **pupil** | **motion** | **pitch** | **rate** | **volume** |
| calm | M | normal | normal | normal | normal |
| fear | M | slow | high | very fast | medium |
| surprise | -?- | very fast | very high | very fast | very loud |
| joy | L | fast | high | fast | loud |
| sorrow | M | very slow | little low | slow | very soft |
| disgust | -X- | normal | low | very slow | soft |
| anger | S | fast | very low | very slow | very loud |

2) Random movements: Random movements are generated. Those affect the gaze direction and speed of movement in function of the robot's mood. The gaze direction tells a lot about the state of mind of human beings. We, therefore, determine a specific window for the random movement in the eye space, which is shown in Fig. 8.

3) Random sequences: For each template expression, a set of movements using eyebrows and eyes can be implemented, e.g., the LED matrix may show a teardrop among other symbols when the robot is sad.

### *B. Mapping Perception to Affects*

The sources taken into account in creating expressions comprise of the following: face tracking; motion detection; buttons; laser scanners; bumpers; and battery. For different conditions, these sources are evaluated with respect to the goals of the robot. The resulting mapping of conditions to desired expressions is shown in Table II. In order to display these expressions, the source information is used to change the internal emotional state, ensuring a smooth transition.

If the robot cannot fullfill its task, it becomes unhappy (sorrowful when nobody is in sight during a presentation; angry if someone plays with the buttons disturbing the robot, or when someone completely blocks the way). The robot is happy when successfully making its job (joyful when seeing someone during a presentation).

### *C. Representation of the Emotional State*

When inputs require the emotional state to change, the expression changes accordingly. It is not credible for all expressions to change instantaneously from, e.g., happy to sad. To do so, we derive a set of intermediate expressions as an interpolation of template expressions, where the transition speed depends on the new emotional state.

We use the three-dimensional (3-D) Arousal–Valence–Stance (AVS) space [44] as an internal representation of the emotional state (see Fig. 9). The advantage of AVS space is that it can be easily mapped to the expression space for the seven template expressions.

Transition in this space results from signals from several sources or explicit scenario inputs, which are transformed to a point of the AVS space $\vec{a}_{\mathrm{input}}$. The new affect $\vec{a}_{\mathrm{new}}$ is computed

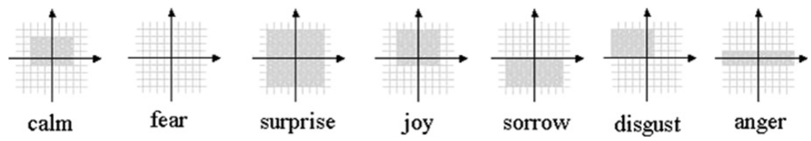


Fig. 8. Parameter range of eye position (pan, tilt) for different template expressions.

TABLE II
SOURCES AND CONDITIONS ORDERED BY PRIORITY WITH THE AFFECT THEY RAISE. EMOTIONAL STATE MACHINE ENSURES SMOOTH TRANSITIONS BETWEEN EXPRESSIONS

| Source | Signal type | Affect |
|---|---|---|
| Battery | low level | sorrow |
| Bumpers | touched front/back | anger |
| Navigation | blocked front/back | anger |
| Buttons | touched without question | anger |
| Motion Detection | nobody in sight | sorrow |
| | $< X$ persons | disgust |
| | $> X$ persons | joy |
| Face Tracking | nobody in sight | sorrow |
| | $< X$ persons | disgust |
| | $> X$ persons | joy |

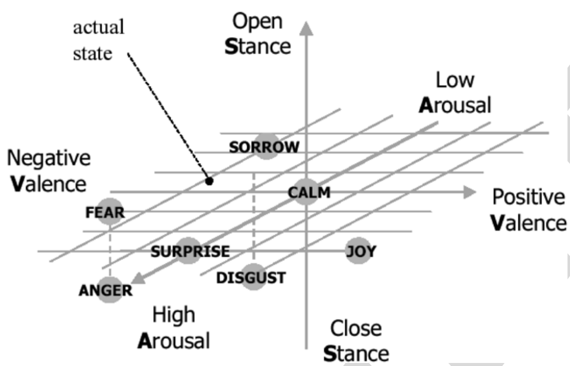


Fig. 9. The robot's emotional state is a point in the AVS space. The robot's seven template expressions are specific states in this space, corresponding to specific output parameters on the expressive modalities. Transitions from one state to another pass through nonmodeled intermediate expressions, which result from interpolation to obtain a smooth transition.

using (1), where $\vec{a}_{\text{prev}}$ denotes the previous affect. The duration of an expression change is denoted by $T$

$$\vec{a}_{\text{new}} = \frac{1}{T+1}(T\vec{a}_{\text{prev}} + \vec{a}_{\text{input}}). \quad (1)$$

The duration of an expression change is a function of the position of the input affect point, particularly of its arousal coefficient. This takes into account the fact that expressions change with different speed. Surprise is usually instantaneous; sorrow, however, comes much slower.

### *D. Expression Generation*

The parameter set $\vec{p}_{\text{new}}$ for the new expression, which is displayed, is a weighted mean of the parameter sets $\vec{p}_e$ for the seven template expressions, denoted as $E$. The inverse of the distance of the current state $\vec{a}_{\text{new}}$ to the template states $\vec{a}_e$ is the weight $w_e$. The new parameter set is given by (2)

$$w_e = (1 + \|\vec{a}_{\text{new}} - \vec{a}_e\|)^{-1}$$
$$\vec{p}_{\text{new}} = \frac{1}{\sum_E w_e} \sum_E w_e \vec{p}_e. \quad (2)$$

Intuitively, the closer the current state is to the center of a template expression, the more the current expression reflects that emotional state. Transitions from one expression to another do not need to be modeled explicitly, but result from the state transition in the affect space as shown in Fig. 10.

## V. INTERACTIVE SCENARIOS

Interactive scenarios are the combination of stage-play presentations and reactive scenarios. By reactive scenarios, we mean small dedicated programs for special situations. Fig. 7 gives an overview of the interactive system.

The scenario composition explains how to create stage-play scenarios for presenting exhibits and reactive scenarios for special situations (robot blocked, battery low). The scenarios may influence the expression directly, by requesting a certain emotional state, or rely on a continuous interpretation of the sensor data to generate expressions.

Stage-play scenarios can combine modalities for interaction (Fig. 11) to create presentations [Fig. 12(a)].

In their simplest form, stage-play scenarios are a linear succession of commands. Introducing parallel execution of tasks increases the scenario's complexity, for instance, allowing to change the facial expression while speaking. Even more complex scenarios contain branches. Such decisions may depend on speech recognition [see the example in Fig. 12(a)], motion detection, or button events.

Two kinds of scenarios are used, namely: 1) presentation scenarios; and 2) reactive scenarios. Depending on the interaction strategy, presentation scenarios are used as a set to create a tour, or dedicated for one application. Presentation scenarios in a tour are executed depending on visitor choices and the availability of free exhibits.

The emotional state machine may inject reactive scenarios into the program, if required, even when a presentation scenario is already running.

When a reactive scenario is triggered, the main program dynamically changes the current presentation scenario. The corresponding reactive scenario is executed until the robot can continue the tour. It is possible to load a number of different scenarios for each case, which allows the robot to vary comments, if the situation did not change after execution of the first reactive scenario.

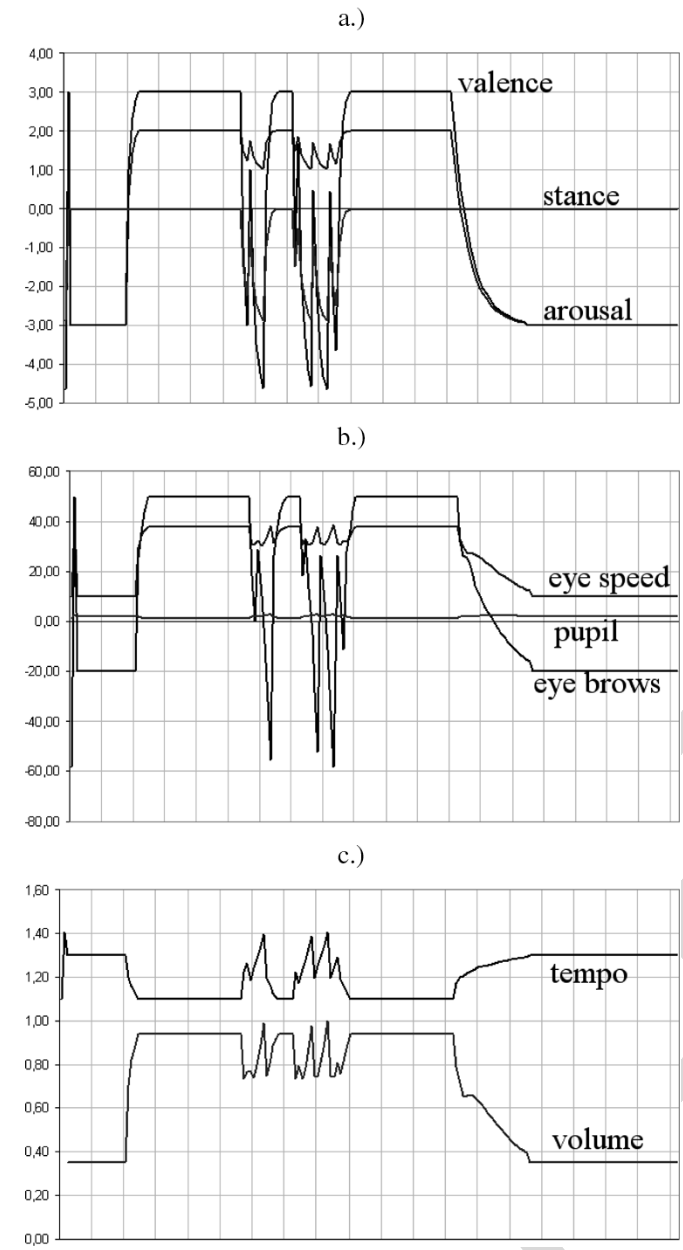


Fig. 10. Relation between affect and expressive modalities during a short experiment. (a) Affect change in the AVS space over time. (b) Parameters for eyes in percent of their maximal value over time. (c) Parameters for synthesized speech, where 1.0 is the default value for volume and speed. In the beginning, nobody is in sight. The robot, thus, shows sorrow until someone arrives. At this time, the arousal value rises very fast, closely following the input arousal signal. The visitor then plays with the buttons, without being asked to use them. The robot becomes nervous and begins to lower its eyebrows. As soon as the visitor stops using the buttons, the joy expression is triggered. Finally, the visitor leaves the robots, which then goes back to a sad expression.

### A. Presentation Scenario

Fig. 12(a) shows a typical presentation scenario. This scenario is executed upon reaching exhibit Alice (F). Assuming people are following the robot, RoboX asks whether or not to present Alice. The answer, given via speech recognition or a button input determines the next step in the scenario. Upon completion of the presentation RoboX continues the tour to a free exhibit.

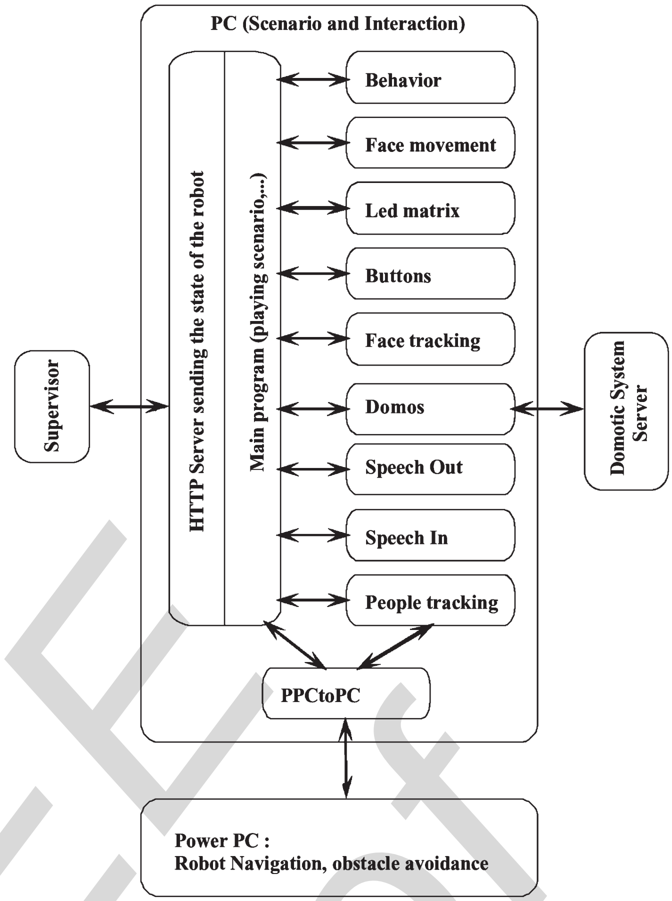


Fig. 11. Block diagram of the main modalities for interaction and how they are linked. Three interfaces function as gateways, namely: 1) the supervision computer; 2) the control of the environment through a dedicated server (Domos); and 3) the navigation part of the robot.

### B. Reactive Scenario

The reaction of RoboX to different situations is programmed with respect to the goals and needs of the tour. For example, if a visitor is blocking the path, RoboX shows anger, because this delays the tour. Cases for which reactive scenarios were developed are as follows: batteries are running low; someone is playing with the buttons; the robot is blocked; and the bumpers are touched. An example is given in Fig. 12(b). It is started when the robot is blocked.

## VI. Results

The exposition *Expo.02* took place from May until October 2002. *Robotics* was one exhibition among several related to different topics. It was open to the public 10 h a day and 12 h during the last month.

The visitors typically spent 10–30 min in the *Robotics@ Expo.02* exhibition. This classifies the man–machine contact as short-term interaction, where the visitors, in contrast to the exposition staff, did not have enough time to form a deeper relationship with the robots as in the experiments reported in [13].

We will report on the overall performance of the robots during the exposition. We try to assess the quality of the

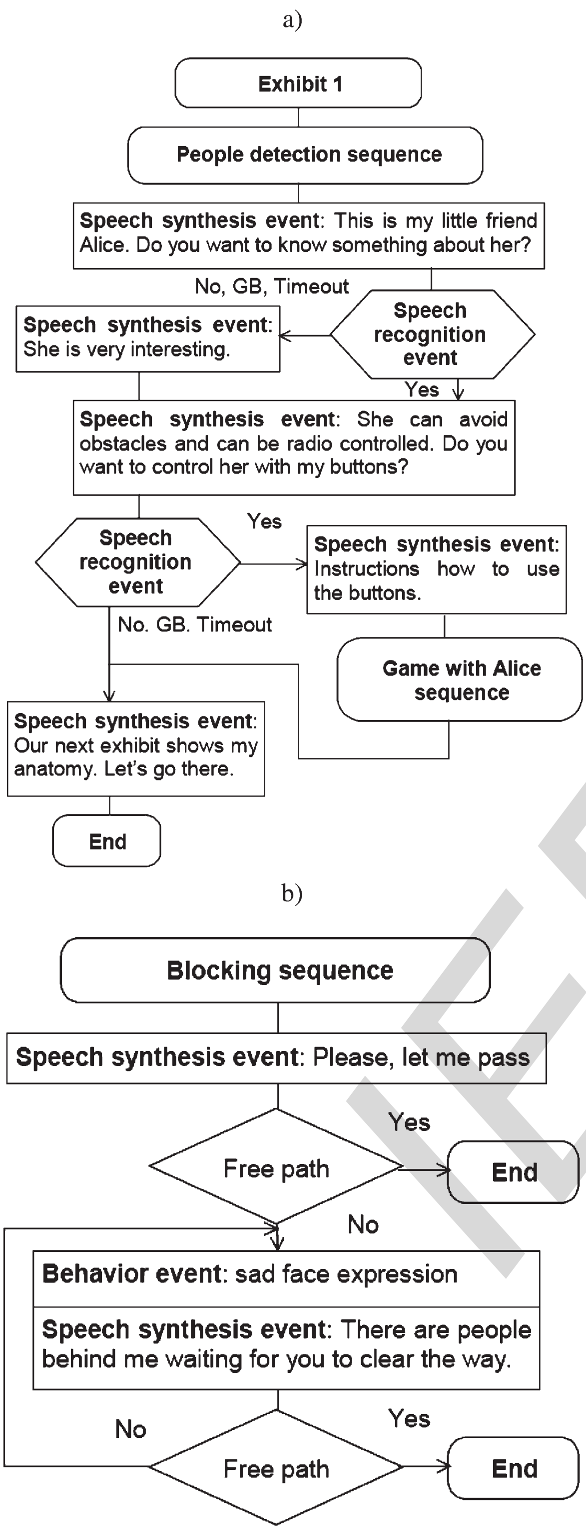


Fig. 12. (a) Sequence presenting the exhibit Alice using people detection, speech synthesis, and recognition. (b) Reactive scenario, which is used when the robot is blocked. When visitors keep RoboX from reaching a goal, it changes its expression. If the obstruction persists, RoboX complains until the way is cleared. In parallel to the scenario, obstacle avoidance tries to circumvent whatever or whoever is blocking the way.

interaction through a survey and analyze the performance of interaction modalities separately.

Throughout the exposition scenarios evolved, presentations changed and new strategies were developed. In conclusion, we

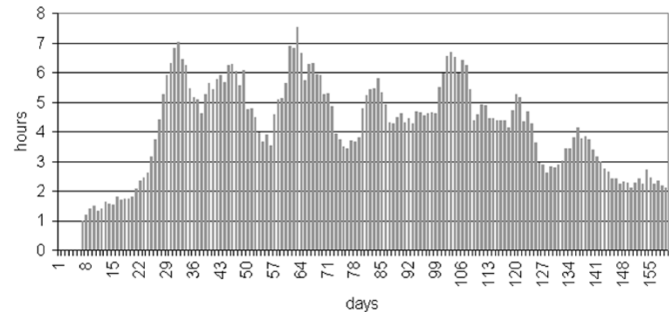


Fig. 13. MTBF as average of 11 robots for each day of the exposition. Note the improvement of MTBF during the first 30 days from 1 to 7 h. During the last month of the exhibition, the MTBF drops again. At the same time, the opening time of the exposition was raised from 10 to 12 h, increasing wear on robots (particularly batteries) and imposing an additional burden on the staff.

report on observations made in the exposition related to these modifications.

### *A. Robot Performance During the Exposition*

During *Expo.02*, 11 RoboXs were guiding more than 686 000 visitors through *Robotics*. Everyday, between 6 and 11 robots were running a 10-h shift each. On the average, 8.4 robots were interacting with 4317 visitors per day (minimum = 2299 and maximum = 5473 visitors), adding up to the following operational values:

1) total run time: 13 313 h;
2) total motion time: 9415 h;
3) traveled distance: 3316 km;
4) maximum speed: 0.6 m/s;
5) average speed: 0.098 m/s;
6) average interactions: 51 visitors/robot/h;
7) mean time between failure (MTBF): 3.26 h.

From the point of view of the performance, MTBF is probably most interesting. Note that a failure is defined as a problem requiring a human intervention in order to allow a robot to continue its work.

Fig. 13 shows the MTBF averaged over 11 robots for each day of the exposition. During the first 30 days, the MTBF increased from 1 to 7 h. This represents the *Robotics@Expo.02* trial phase. Despite our demands, on-site testing prior to the begin of the exposition to was limited to two days.

During the last month of the exhibition, the MTBF drops again. One reason for this is the extension of the opening time from 10 h, for which the robot were designed, to 12 h. It not only increased the wear on the robots, particularly the batteries, but also imposed an additional burden on the staff. Consequently, visitors were not always stopped when abusing the robots by kicking or pushing them around. A detailed analysis of performance data can be found in [45].

Summarizing, we judge the MTBF of 3.26 h per robot as satisfactory for a system built from scratch within a year. This MTBF corresponds to approximately 25 human interventions per day for the whole exhibition.

Regarding the safety aspects, we neither received complaints nor did we observe any dangerous situations. Accidents did not occur. When not obstructed intentionally by visitors, obstacle avoidance was able to guide RoboX, even in tight situations without collision. Of course, intentional obstructions occurred. The low speed of RoboX and its immediate stopping on contact made blocking the robot's way a popular and harmless game for visitors.

### *B. Results From Survey*

We made a survey to evaluate the quality of the exposition and the importance of the different modalities. The queried visitor had to answer the following questions:

1) How do you rate the robot's appearance?
2) How do you rate the robot's character?
3) How good is the synthesized speech?
4) How did you learn to use the robot?
5) How do you rate the speech recognition? (only on two robots)
6) Which sensor is used for navigation?
7) Which exhibits did you visit?
8) How do you rate the exhibition?
9) Would you prefer a normal information desk or an interactive robot when asking for directions?

Answers were collected from 209 visitors, 106 (58%) female and 89 (42%) male, speaking German 128 (61%) , French 75 (36%), or Italian 6 (3%). The average age was 34.4 years, the oldest participant was 74 years old, and the youngest was five years old.

The aggregated results to questions 1, 2, and 8 show a very similar distribution as follows: very good (20%); good (51%); acceptable (26%); bad (3%) within a small margin (3%). This strongly suggests that, during the short time of their stay, visitors perceived the robots, probably the entire exposition as a whole.

Speech synthesis (question 3) was rated above the overall average with a distribution as follows: very good (31%); good (44%); satisfactory (24%); and bad (1%). The same applies for speech recognition (question 5) with a distribution as follows: very good (37%); good (39%); satisfactory (20%); and bad (4%).

When asked how they learned to use the robot (question 4), most visitors selected the first answer (from the robot itself), as shown in Fig. 14(a). However, the fact that 11% did not learn to use the robots shows that the reluctance to touch and interact with a machine is not negligible and particular effort has to be made to ease the first contact.

In the same survey, visitors were asked questions about the functioning of the robot (question 6). As shown in Fig. 14(b), more than two thirds of the visitors understood that robots use laser sensors and not eyes for navigation.

These results probably explain why the visitors would prefer the robot (72%) to an information desk (28%) to ask for directions (question 9) in places like train stations or expositions.

### *C. Evaluation of Modalities for Interaction*

Regarding the modalities for interaction, we were interested in the reliability of motion detection, face tracking, and speech recognition under *Expo.02* conditions. Concerning the

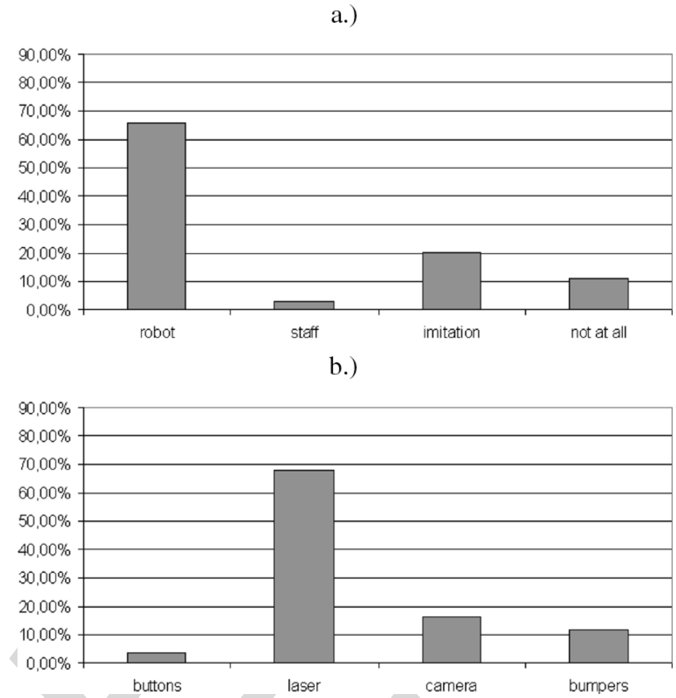


Fig. 14. Results from the survey. Only one selection was possible. (a) How did the visitors learn how to use the robot? The answers from the visitors show that the robot itself was the best teacher. Note that only 11% of the visitors did not learn how to use the robot. (b) Understanding of elementary principles taught by the tour-guide robot. Two hundred nine visitors have been asked to say what was the main sensor used for navigation. More than two thirds understood correctly that it was the laser.

TABLE III
EXPERIMENTAL RESULTS FOR MOTION DETECTION FOR A SEQUENCE OF 279 SCANS

| | scans | present | detected | error (I) | error (II) |
|---|---|---|---|---|---|
| **total** | 279 | 2461 | 2289 | 238 | 66 |
| **average** | | | 90.9% | 9.2% | 2.8% |

expressive modalities, we wanted to know whether visitors could understand the synthesized speech and the expressions generated.

To evaluate the perceptive modalities, we manually evaluated sequences from *Expo.02* and compared this to the results that RoboX obtained. The testing terminology is as follows: By detected, we refer to all those elements that were correctly detected. The detection rate is the ratio of correct recognition to all correct elements. A type-I error is the rejection of a correct element; it refers to the number of correct elements present. Finally, a type-II error is the failure to reject a wrong element; it relates to the sum of correct and false detection.

*1) Motion Detection:* Motion detection was evaluated on a sequence of 279 scans from the robot Photographer (L). The number of persons visible, the number of persons not detected as a motion cluster, and the number of clusters not corresponding to a person were counted for each scan. Persons not visible in the scan due to occlusion were not considered. Table III summarizes the results.

On the average, nine persons were present in a scan. The minimum was 5 and the maximum was 14 persons. The type-I

TABLE IV
EXPERIMENTAL RESULTS OF FACE TRACKING, FROM AN 11-MIN SEQUENCE. EVALUATION LIMITED TO 169 IMAGES (EVERY TWENTIETH) DUE TO SIMILARITY OF SUCCESSIVE IMAGES

| image type | images | present | detected | error (I) | error (II) |
|---|---|---|---|---|---|
| sharp | 100 | 584 | 375 | 246 | 37 |
| blurred | 39 | 193 | 88 | 105 | 0 |
| dark | 30 | 270 | 34 | 236 | 0 |
| **total** | 169 | 1047 | 497 | 587 | 37 |
| sharp | | | 64.2% | 42.1% | 8.9% |
| blurred | | | 45.6% | 54.4% | 0.0% |
| dark | | | 12.6% | 87.4% | 0.0% |
| **average** | | | 47.5% | 56.1% | 6.9% |

error was found to increase with the number of persons present. Dense crowds of visitors often caused partial occlusions. The remaining motion clusters were too small to be considered as a person and accumulated to an error of 9.2%.

Regarding the environment, Photographer (L) was operating in a very structured part of *Robotics@Expo.02*. Different from those robots operating in the main hall, a high percentage of its scans represented static environment. Despite this, static elements were rarely confused with motion. The error remained small 2.8%. The overall detection rate for motion amounts to 90.9%.

*2) Face Tracking:* The performance of the face tracking algorithm was evaluated quantitatively from a sequence of images, similar to the one shown in Table IV. The sequence lasting 11 min was sampled at 4 Hz resulting in 2800 images. The manual evaluation of the faces present, detected and tracked per image, was limited to every twentieth image, since consecutive images are very similar. In total, 169 images were classified. The results are summarized in Table IV. Images were classified in categories. We distinguish images as follows: sharp images; images with motion blur; and dark images. The dark image class comprises a part at the beginning of the sequence with very low illumination, for which the skin color model was not designed.

At the beginning of the sequence, a robot welcomes a group of visitors. Here, on the average, there were nine faces in the images, whereas in the remainder of the sequence, the average number drops to five or six faces.

In the 169 images evaluated, a total of 1047 faces were present, of which 497 were correctly detected. A total of 37 regions were detected, which did not correspond to a face, resulting in a type-II error of 6.9%. The detection rate was 47.5% on the average and 64.2% for sharp images. The detection rate drops to 12.59% for dark images. This is probably due to the skin color model, which was created for normal illumination.

For motion detection, the type-I error increases again with the number of persons present, probably due to partial occlusions. The detection rate of 47.5% (64.2% sharp images) is in part due to the crowded situation of up to 11 faces on the images, which cover a considerable smaller angle than the laser sensors. The type-II error is still low (8.9%), so that RoboX almost never assumed the presence of a person, when, in fact, there was none.

TABLE V
EXPERIMENTAL RESULTS FOR SPEECH RECOGNITION. RECOGNITION OF 130 TEST SAMPLES FROM *Expo.02* FOR THE GARBAGE MODEL, YES AND NO EACH. COMPARISON OF RESULTS FROM OBSERVED RECOGNITION RESULTS OF PLAIN SPEECH RECOGNITION (ORR) AND BAYESIAN NETWORKS (BNS) FUSING SPEECH RECOGNITION AND LASER DATA

| | garbage model | yes | no | detection |
|---|---|---|---|---|
| ORR Acc | 38.5% | 93.1% | 66.9% | 66.2% |
| BN Acc | 80.8% | 84.6% | 66.9% | 77.4% |
| **Gain** | 42.3% | -8.5% | 0.0% | 11.3% |

*3) Speech Recognition:* After the *Expo.02*, additional experiments were made to overcome the recognition errors in noisy conditions. We found that combining the speech recognition result with additional information from acoustic noise-insensitive laser scanner data can lead to improved speech recognition performance.

In Table V, results from plain speech recognition (ORR) are compared to the new BN-based approach. This is explained in detail in [46].

The results show that the original system achieved good recognition results for yes (93.1%) and no (66.9%), but suffered from a weak detection for the garbage model. Fusing the recognition results with laser scanner data improved the detection (80.8%). Sometimes, laser data indicated the absence of persons, when, in fact, they were present and answering, this explains why the BN recognition result for yes drops to 84.6%.

*4) Synthesized Speech:* As found in the survey (Section VI-B), visitors rated the quality of the synthesized speech even above the overall exposition impression. This is further supported by discussions with visitors, where we learned that the quality of synthesized speech was different for each language. Synthesized French was understandable, English and German were found to be good, and Italian even excellent.

We would like to raise attention to the point that people sometimes mentioned the recording of the speaker could have been better and were surprised to learn that there was no natural speech involved at all. Here, it appears as if the robot came to close to imitate our natural speech, thus, raising visitor expectation from communicating with a machine to the variations in pronunciation a professional speaker delivers.

*5) Expressions:* In the context of an exhibition, visitors expect surprise and something out of the ordinary. This creates a certain liberty regarding the appearance of the robot. To create expressions, RoboX even used an asymmetric mechanical face without a mouth. Even if the visitor is prepared for something unusual, the template expressions should be readily discernable (Fig. 15).

Prior to *Expo.02*, we tested the recognition with a group of 37 test persons. The results in Table VI show that fear, sorrow, and joy where well recognized. Disgust, anger, and surprise show poor results.

Apparently, recognition of the latter three expressions relies on the shape of the mouth. Consequently, for *Expo.02*, we included symbols for the different expressions. Fig. 6 shows

Fig. 15. Photobot (L) in its booth taking pictures of visitors. Selected photos: how people react to the robot photographer. The final image shows the Cadavre Exquis (N), where recently taken photos were shown by mixing parts of visitor photos with robot parts, creating artificial cyborgs.

TABLE VI
EXPERIMENTAL RESULTS FOR RECOGNITION OF FACIAL EXPRESSIONS. PERCENT OF CORRECTLY RECOGNIZED EXPRESSIONS FROM A GROUP OF 37 PERSONS IS SHOWN

| fear | sorrow | joy | disgust | anger | surprise |
|---|---|---|---|---|---|
| 55.0% | 85.7% | 55.4% | 0.0% | 5.4% | 12.5% |

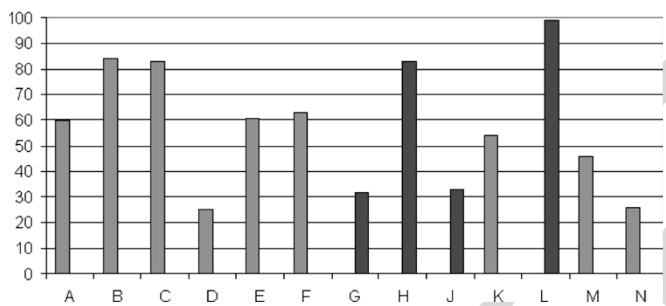


Fig. 16. Number of visitors per exhibit. Exhibits are arranged according to their distance from the entry. Dark bars indicated the robots as exhibits and lighter bars indicate the tour-guide exhibits. The corresponding locations are shown in Fig. 1. There are strong variations between both groups. It is interesting to note that with Medical robot (B) and Me, myself and I (C), the first stations of the tour are the most crowded. The Photobot (L) and Jukebot (J) succeed in attracting visitors even toward the exit of the exhibition. The location of less popular stations (D,G, J) is between the wall and the bioscope, which was outside the mainstream of visitors. The first tour-station Industry robot (A) and the last Cadavre Exquis receive less visitors due to effects of forming groups and leaving the exposition.

the use of a question mark for surprise and an X-symbol for disgust, creating more distinctive expressions.

## VII. DISCUSSION

The discussion comprises an assessment of interaction strategy by means of visitor density, a report on the evolution of scenarios and changes in the exhibition, and personal impressions from staff members, who worked in *Robotics@Expo.02* throughout the 5-month period.

### A. Interaction Strategies and Visitor Density

In the survey, visitors were asked which stations the robot presented to them. The distribution is shown in Fig. 16. Labels correspond to locations in Fig. 1. Exhibits are ordered according to their distance from the entry.

As was pointed out earlier, visitors perceived the exhibition as a whole, making it difficult to evaluate different types of interaction directly with a survey. However, visitors correctly remembered which part of the exposition they visited. We argue that the number of visitors per exhibit indicates its popularity and try to infer from this which types of interactions were appealing to visitors.

Particular interest received: Photobot (L) and Jukebot (J), which were not part of the guided tour, but were served by a dedicated RoboX. Among the tour stations, two of the three foremost stations received the most attention [Medical robot (B) and Me, myself and I (C)].

Visitors started the exhibition by joining a guided tour provided by the robots. With the exception of Fossil (D), the number of persons per guided group decreased gradually toward the exit, probably because they were attracted to other parts of the exhibition. Our observations throughout *Expo.02* confirm the visitor distribution derived from the survey and shown in Fig. 16. In our opinion, the lack of visitors at Industrial robot (A) was due to its proximity to the welcome area. Visitors sometimes started tours inadvertently, selecting the wrong language. Instead of following the robot, they joined another tour in their language given by one of the other robots nearby. In fact, when we moved the welcome area from around point (A) into the hallway near point (Z), more visitors were attracted to Industrial robot.

The Fossil (D) exhibit was presented using the same techniques as Medical robot (B), Me, myself and I (C), and Aquaroids (E). The lack in visitors may be attributed to its location as it is not in the exhibition's mainstream. This may as well apply to the Presenter robot (G) located nearby, which was explaining some insights of RoboX using projected slides. Stations that explained robot perception were Face Tracking (K) and Supervision Lab (M).

The noticeable interest in the exhibits Photobot (L) and Jukebot (H) convinced us that short and highly reactive scenarios create an interesting interaction for the visitor, since their actions were immediately rewarded by the robot.

### B. Scenario Evolution

Stage-play scenarios were revised throughout *Expo.02*, reflecting experience gathered during the exhibition. As an example of this evolution, the introduction scenario is outlined. Then we address the issue of timing with regards to visitor behavior and robot reaction.

*1) Introduction Scenario:* A critical point in the exposition was the first contact of visitors and robots. The problem was explaining how to operate the robot to select the tour language, without knowing the visitor's language. In case of selecting the wrong language, visitors normally ceased interaction with this robot and moved on to another.

The introduction scenario was revised several times. Two independent versions were maintained, one for the two robots with speech recognition and one for those using buttons only.

In the first versions of the voice-enabled introduction scenario, RoboX asked four questions, "Do you speak English/German/French/Italian?" in the four official languages. Although these questions implied a yes/no answer, people often expected the robot to understand utterances such as "No Italiano" or "Ich spreche Deutsch." To avoid this, we refined the questions to: "For English/French/German/Italian, answer with yes/oui/ja/si or no/non/nein/no" in the four languages supported by the interface. This made the "introduction sequence" longer than before, but more effective.

Similar problems arose for introduction scenario using buttons. It started with the question sequence "red—French/blue—German/green—English/orange—Italian". When saying "red for French," some visitors immediately pressed on the red alarm button instead of waiting for the end of the sentence and choosing by pressing on the red colored button.

The best working solution for the introduction scenario finally consisted in attracting interest using an artificial babble language, explaining the language choice in all four languages, confirming the choice, and eventually starting the tour.

Moving the place where robots were waiting for the visitors from the main hall [around point (A)] into the hallway [close to (Z)] resulted in a more reliable language selection. Here, visitors were not yet confronted with the entire exhibition and could better focus on one robot, reducing the problem of false language selection.

*2) Timing:* In the context of questions and answers, as in the combination of stage-play and reactive behavior, timing was found to be of particular importance.

When initially creating scenarios, we expected the robot to state a question and then visitors to answer during a certain lapse of time. However, in reality, visitors had a tendency to reply immediately, even before the robot finished the question and was prepared to handle the answer. Other visitors hesitated or were undecided until the robot quit expecting an answer.

This was particularly difficult for speech recognition. The noisy conditions in the first case lead to recognition errors. The failure to act correctly upon answers lead to disappointment. Thus, as an additional information, the LED matrix display was used to signal the right moments for answering using start and stop symbols. In the case of button input, flashing lights around the buttons were used to indicate when the robot was waiting for an answer.

Timing was also found to be an issue when combining stage-play and reactive scenarios. Sometimes, events like touching the buttons occurred while the robot was in the middle of a long task; when it finally responded to the event after task completion, the situation sometimes had evolved so much, so that the relation of event and scenario was difficult to discern for the inexperienced visitor. As a remedy to enable faster reaction, robot speech was changed from long monologues to short phrases.

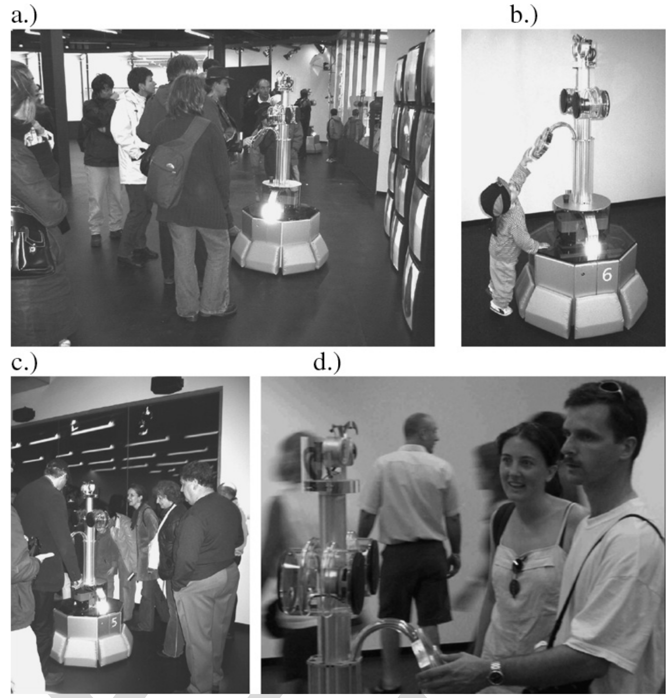


Fig. 17. Some impressions from *Robotics@Expo.02*. Visitors interacting with RoboX. (a) Group of visitors in front of Cadavre Exquis (N). In the background is the Photographer (L). (b) Child stretching for buttons. (c) Group of visitors near Industry robot (A). (d) Couple selecting next tour station.

## C. Impressions

From discussion and observation of the exposition, we learned that visitors appreciate robots that react quickly and in a diverse nonforeseeable way. This is further confirmed by the success of reactive scenarios with visitors and their enthusiasm in playing with the obstacle avoidance. Blocking the way, touching buttons, or kicking bumpers rarely ceased after complaints from the robot. On the contrary, our efforts in making complaints vary only increased visitors persistence (Fig. 17). AQ5

From a system design perspective, reactive scenarios are needed to support the robot in reaching its goals more quickly. From an interaction point of view, we judge their extensive use by visitors as a success.

When trying to get RoboX attention, visitors were often seen waving hands in front of its mechanical face. We see this as acceptance of the face as an anchor of communication, supporting the concept of a mechanical yet familiar face.

Regarding the attachment to the robot, it is interesting to compare the visitor's behavior to that of the exposition staff. As mentioned earlier, visitors perceived the exposition as a whole, whereas staff was referring to each RoboX individually, assigning it a particular character based on its individual operational performance.

Visitors were willing to learn how to interact. Children particularly seemed to understand the robot easily in their playful manner. Sometimes, visitors' curiosity went beyond limits, as in the case of the alarm button. Originally intended as a safety feature, it stopped the robot immediately and activated an alarm sound. This unintentionally made it a popular feature among some visitors.

## VIII. Conclusion

This paper has presented experiences of a long-term exhibition *Robotics@Expo.02* with 11 mobile robot tour guides. The design and implementation of the tour-guide robot (RoboX) have been described. Aspects of reliability and safety in public space have been addressed, and human–robot interaction during the exhibition has been assessed.

The objectives of interaction, the exhibition, and its development have been presented. Robotic modalities for interaction have been presented in detail. Perceptive elements (motion detection, face tracking, speech recognition, buttons) have been distinguished from expressive ones (robotic face, speech synthesis, colored button lights). An approach for combining stage-play and reactive scenarios has been presented. An emotional state machine has been used to create convincing expressions from the robot.

For the entire 5-month duration of the exhibition, an evaluation of the robot performance has been given. A performance analysis of modalities for interaction has also been presented. Survey results to assess human–robot interaction and interaction strategies have also been included.

The event *Robotics@Expo.02* has greatly contributed to our experience in the field of large-scale human–robot interaction. We hope that the results will contribute to the further development of interactive robots.

## Acknowledgment

The production of the 11 RoboXs was realized by BlueBotics, a spin-off of the Autonomous Systems Lab. The authors thank all members of the team *Robotics@Expo.02* for their outstanding contributions, namely: K O. Arras; M. de Battista; S. Bouabdallah; D. Burnier; G. Froidevaux; X. Greppin; B. Jensen; A. Lorotte; L. Mayor; M. Meisser; R. Philippsen; P. Prodanov; R. Piguet; G. Ramel; M. Schild; R. Siegwart; G. Terrien; and N. Tomatis. Apart from this core team, various people from academia and industry supported the project. The authors are particularly grateful to R. Philippsen for his help in preparing the paper. The authors also thank P. Prodanov for sharing his expertise on speech recognition and S. Vasudevan for fruitful discussions.

## References

[1] M. Fujita, "AIBO: Toward the era of digital creatures," *Int. J. Rob. Res.*, vol. 20, no. 10, pp. 781–794, Oct. 2001.

[2] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent ASIMO: System overview and integration," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Lausanne, Switzerland, 2002, vol. 3, pp. 2478–2483.

[3] *The McGraw–Hill Illustrated Encyclopedia of Robotics and Artificial Intelligence*, McGraw-Hill, New York, Jan. 1995.

[4] C. Breazeal, "A motivational system for regulating human–robot interaction," in *Proc. Amer. Association Artificial Intelligence*, Madison, WI, 1998, pp. 54–61.

[5] B. 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," *Artif. Intell.*, vol. 114, no. 1–2, pp. 3–55, Oct. 1999.

[6] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Ahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "Probabilistic algorithms and the interactive museum tour-guide robot Minerva," *Int. J. Rob. Res.*, vol. 19, no. 11, pp. 972–999, Nov. 2000.

[7] B. Graf, R. Schraft, and J. Neugebauer, "A mobile robot platform for assistance and entertainment," in *Proc. 31st Int. Symp. Robotics*, Montreal, Canada, 2000, pp. 252–253.

[8] I. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto, "An effective mobile robot educator with a full-time job," *Artif. Intell.*, vol. 114, no. 1–2, pp. 95–124, Oct. 1999.

[9] T. Willeke, C. Kunz, and I. Nourbakhsh, "The history of the Mobot museum robot series: An evolutionary study," in *Proc. Florida Artificial Intelligence Research Society (FLAIRS)*, Key West, FL, 2001, pp. 514–518.

[10] R. Bischoff and V. Graefe, "Demonstrating the humanoid robot HERMES at an exhibition: A long term dependability test," in *Workshop: Robots Exhibitions, IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Lausanne, Switzerland, 2002. **AQ6**

[11] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human–robot interaction," in *Proc. Amer. Association Artificial Intelligence (AAAI) Fall Symp.*, Boston, MA, 2001. **AQ7**

[12] ——, "The role of expressiveness and attention in human–robot interaction," in *Proc. IEEE Int. Conf. Robotics and Automation*, Washington, DC, 2002, pp. 4138–4142.

[13] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "A practical experiment with interactive humanoid robots in a human society," in *IEEE Int. Conf. Humanoid Robots*, Munich, Germany, 2003. **AQ8**

[14] *Robotics@Expo.02.* [Online]. Available: http://robotics.epfl.ch **AQ9**

[15] K. O. Arras, J. A. Castellanos, and R. Siegwart, "Feature-based multi-hypothesis localization and tracking for mobile robots using geometric constraints," in *IEEE Int. Conf. Robotics and Automation*, Washington, DC, 2002, pp. 1371–1377.

[16] O. Brock and O. Khatib, "High-speed navigation using the global dynamic window approach," in *Proc. IEEE Int. Conf. Robotics and Automation*, Detroit, MI, 1999, pp. 341–346.

[17] J.-C. Latombe, *Robot Motion Planning*. Dordrecht, The Netherlands: Kluwer, 1991.

[18] S. Quinlan and O. Khatib, "Elastic bands: Connecting path planning and control," in *Proc. IEEE Int. Conf. Robotics and Automation*, Atlanta, GA, 1993, pp. 802–807.

[19] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robot. Autom. Mag.*, vol. 4, no. 1, pp. 23–33, Mar. 1997.

[20] C. Schlegel, "Fast local obstacle avoidance under kinematic and dynamic constraints for a mobile robot," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Victoria, Canada, 1998, pp. 594–599.

[21] R. Philippsen and R. Siegwart, "Smooth and efficient obstacle avoidance for a tour guide robot," in *Proc. IEEE Int. Conf. Robotics and Automation*, Taipei, Taiwan, 2003, pp. 446–451.

[22] R. Brega, N. Tomatis, K. Arras, and R. Siegwart, "The need for autonomy and real-time in mobile robotics: A case study of xo/2 and pygmalion," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Takamatsu, Japan, 2000, pp. 1422–1427.

[23] P. Prodanov, A. Drygajlo, G. Ramel, M. Meisser, and R. Siegwart, "Voice enabled interface for interactive tour guide robots," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Lausanne, Switzerland, 2002, pp. 1332–1337.

[24] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "MINERVA: A second-generation museum tour-guide robot," in *Proc. Int. Conf. Robotics and Automation (ICRA)*, Detroit, MI, 1999, vol. 3, pp. 1999–2005.

[25] E. Prassler, J. Scholz, and E. Elfes, "Tracking people in a railway station during rush-hour," in *Proc. Int. Conf. Computer Vision Systems (ICVS)*, Las Palmas, Spain, 1999, pp. 162–179.

[26] ——, "Tracking multiple moving objects for real-time robot navigation," *Auton. Robots*, vol. 8, no. 2, pp. 105–116, Apr. 2000.

[27] D. Schulz, W. Burgard, D. Fox, and A. Cremers, "Tracking multiple moving targets with a mobile robot using particle filters and statistical data

association," in *Proc. IEEE Int. Conf. Robotics and Automation*, Seoul, Korea, 2001, pp. 1665–1670.

[28] B. Jensen and R. Siegwart, "Using EM to detect motion with mobile platforms," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Las Vegas, NV, 2003, pp. 1518–1523.

[29] B. Jensen, R. Philippsen, and R. Siegwart, "Narrative situation assessment for human–robot interaction," in *Proc. IEEE Int. Conf. Robotics and Automation*, Taipei, Taiwan, 2003, pp. 1503–1508.

[30] B. Jensen, G. Froidevaux, X. Greppin, A. Lorotte, L. Mayor, M. Meisser, G. Ramel, and R. Siegwart, "Multi-robot–human interation and visitor flow management," in *Proc. IEEE Int. Conf. Robotics and Automation*, Taipei, Taiwan, 2003, pp. 2388–2393.

[31] A. Hilti, B. Nourbakhsh, I. Jensen, and R. Siegwart, "Narrative-level visual interpretation of human motion for human–robot interaction," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Maui, HI, 2001, pp. 2074–2079.

[32] B. Jensen, G. Froidevaux, X. Greppin, A. Lorotte, L. Mayor, M. Meisser, G. Ramel, and R. Siegwart, "The interactive autonomous mobile system robox," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Lausanne, Switzerland, 2002, pp. 1221–1227.

[33] A. Drygajlo, P. Prodanov, G. Ramel, M. Meisser, and R. Siegwart, "On developing voice enabled interface for interactive tour-guide robots," *J. Adv. Robot., Robot. Soc. Jpn.*, vol. 17, no. 7, pp. 599–616, Nov. 2003.

[34] D. Spiliotopoulos, I. Androutsopoulos, and C. Spyropoulos, "Human–robot interaction based on spoken natural language dialogue," in *Eur. Workshop Service and Humanoid Robots (ServiceRob)*, Santorini, Greece, 2001, pp. 1057–1060.

[35] S. Young, J. Odell, D. Ollason, and P. Woodland, *The HTK Book, Version 3.0*. Redmond, WA: Microsoft Corp., 2000.

[36] P. Renevey and A. Drygajlo, "Securized flexible vocabulary voice messaging system on Unix workstation with ISDN connection," in *Eur. Conf. Speech Communication and Technology (Eurospeech)*, Rhodes, Greece, 1997, pp. 1615–1619.

[37] X. Huang, A. Acero, and H. Hon, *Spoken Language Processing*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[38] C. Breazeal and B. Scasselatti, "How to build robots that make friends and influence people," in *Proc. IEEE Int. Conf. Intelligent Robots and Systems*, Kyonjiu, Korea, 1999, pp. 858–863.

[39] B. Siebenhaar-Rölli, B. Zellner-Keller, and E. Keller, *Phonetic and Timing Considerations in a Swiss High German TTS System*. New York: Wiley, 2001, pp. 165–175.

[40] E. Keller and B. Zellner, *A Timing Model for Fast French*. York, U.K.: Univ. York, 1996, pp. 53–75.

[41] T. Dutoit, V. Pagel, N. Pierret, F. Bataille, and O. V. der Vrecken, "The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes," in *Proc. Int. Conf. Spoken Language Processing (ICSLP)*, Philadelphia, PA, 1996, vol. 3, pp. 1393–1396.

AQ10 [42] *IBM ViaVoice* [Online]. Available: http://www-306.ibm.com/software/voice/viavoice/

[43] L. Mayor, B. Jensen, A. Lorotte, and R. Siegwart, "Improving the expressiveness of mobile robots," in *Proc. Robot and Human Interactive Communication (ROMAN)*, Berlin, Germany, 2002, pp. 325–330.

[44] P. Ekman and R. Davidson, *The Nature of Emotion: Fundamental Questions*. New York: Oxford Univ. Press, 1994.

[45] N. Tomatis, G. Térrien, R. Piguet, D. Burnier, S. Bouabdallah, and R. Siegwart, "Designing a secure and robust mobile interacting robot for the long term," in *IEEE Int. Conf. Robotics and Automation*, Taipei, Taiwan, 2003, pp. 4246–4251.

[46] P. Prodanov and A. Drygajlo, "Bayesian networks for spoken dialogue management in multimodal systems of tour-guide robots," in *Proc. 8th Eur. Conf. Speech Communication and Technology*, Geneva, Switzerland, 2003, pp. 1057–1060.

AQ11 AQ12 **Björn Jensen** (S'02–M'04) received the master's degree in electrical engineering and business administration from the Technical University of Darmstadt, Germany, in 1999. He is working toward the Ph.D. degree at the Autonomous Systems Lab (ASL), Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

His main interest is in enhancing man–machine communication using probabilistic algorithms for feature extraction, data association, tracking, and scene interpretation.

**Nicola Tomatis** received the M.Sc. degree in computer science from the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, in 1998, and the Ph.D. degree from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 2001. AQ13

His research covered metric and topological (hybrid) mobile robot navigation, computer vision, and sensor data fusion. Since autumn 2001, he holds a part-time position as Senior Researcher with the Autonomous Systems Lab. He is currently the CEO of BlueBotics SA, Lausanne, Switzerland, which is a start-up involved in mobile robotics.

**Laetitia Mayor** studied at EPFL and Carnegie Mellon University and received the master's degree in microengineering from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 2002. In her master's thesis, she developed a concept for emotional human–robot interaction. AQ14

In spring 2002, she joined the *Expo.02* robotics team at EPFL to work on emotional human–robot interaction and the development of scenarios. After the successful completion of the *Expo.02* project, she joined Helbling Technik AG.

**Andrzej Drygajlo** (M'84) received the M.Sc. and Ph.D. (*summa cum laude*) degrees in electronics engineering from the Silesian Technical University, Gliwice, Poland, in 1974 and 1983, respectively.

In 1974, he joined the Institute of Electronics at the Silesian Technical University where he was an Assistant Professor from 1983 to 1990. Since 1990, he has been affiliated with the Signal Processing Laboratory (LTS) of the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, where he presently works as a Research Associate. In 1993, he created the Speech Processing Group of the LTS. His current research interests are man–machine communication, speech processing, and biometrics. Currently, he conducts research and teaching in these domains at the EPFL and the University of Lausanne. He participates in numerous national and international projects and is member of various scientifc committees. He is currently an Advisor on numerous Ph.D. theses. He is the author/coauthor of more than 70 research publications, including several book chapters, together with his own book publications. He is also an appointed expert nominated by the European Commission in the domain of speech and language technology.

Dr. Drygajlo is a member of the EURASIP, International Speech Communication Association (ISCA), and European Circuit Society (ECS) professional groups.

**Roland Siegwart** (M'90–SM'03) received the M.Sc. degree in ME and the doctoral degree from the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, in 1983 and 1989, respectively. AQ16

After his Ph.D. studies, he spent one year as a postdoc at Stanford University, where he was involved in microrobots and tactile gripping. From 1991 to 1996, he worked part time as R&D Director at MECOS Traxler AG and as a Lecturer and Deputy Head at the Institute of Robotics, ETH. Since 1996, he has been a Full Professor for Autonomous Systems and Robots at the Swiss Federal Institute of Technology, Lausanne (EPFL), and since 2002, a Vice Dean of the School of Engineering. He leads a research group of around 25 people working in the field of robotics and mechatronics. He has published over 100 papers in the field of mechatronics and robotics, is an active member of various scientific committees, and is a cofounder of several spin-off companies.

Dr. Siegwart was the General Chair of IROS 2002 and is currently VP for Technical Activities of the IEEE Robotics and Automation Society.

# AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ1 = Specify location (i.e., city, zip code, and country).
AQ2 = Note that Fig. 7, which was not directly cited in the text, was inserted here.
AQ3 = Note also that 4 was changed to Table 4 (callout). Is this correct?
AQ4 = Note that Fig. 15, which was not directly cited in the text, was inserted here.
AQ5 = Note that Fig. 17, which was not directly cited in the text, was inserted here.
AQ6 = Please provide page range in Ref. [10].
AQ7 = Please provide page range in Ref. [11].
AQ8 = Please provide page range in Ref. [13].
AQ9 = Please provide additional information, if possible in Ref. [14].
AQ10 = Please provide additional information, if possible in Ref. [42].
AQ11 = Specify the type of degree.
AQ12 = Indicate the major field of study.
AQ13 = Indicate the major field of study.
AQ14 = Specify the type of degree.
AQ15 = Is this the M.Sc. degree in mechanical engineering?
AQ16 = Specify the type of degree as well as the major field of study.
Note: Figures 1, 3, 5, 10, 13–16 were processed as grayscale/B&W.

END OF ALL QUERIES