

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).

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].

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

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

103 The Mobot Museum Robot Series reported in [8] and [9]
104 puts more focus on interaction and design, simplifying its
105 navigation task by means of artificial landmarks in the envi-
106 ronment. The robots Sage [8], Chips, Sweetlips, Joe, and Adam
107 [9] emerged over the years and used an increasing number of
108 interaction modalities. They operated for up to three years with
109 Sage covering a total of 323 km [8] and Chips, Sweetlips,
110 Joe, and Adam each covering more than 600 km [9]. With
111 the exception of the last mentioned robot, the movements of
112 the others were limited to a predefined set of unidirectional
113 safe routes in order to simplify both localization and path
114 planning.

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

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

133 C. *Expo.02*

134 The Swiss National Exhibition takes place approximately
135 once every 40 years. *Expo.02* took place from May 15 to
136 October 21, 2002. It was a major national happening with
137 37 exhibitions and an event-rich program. The *Robotics@*
138 *Expo.02* exhibition [14] was intended to show the increas-
139 ing closeness between humans and robots. The central visitor
140 experience of *Robotics@Expo.02* was the interaction with au-
141 tonomous freely navigating mobile robots giving guided tours
142 and presenting the exhibits shown in Fig. 1. The exhibition
143 was scheduled for a visitor flow of 500 persons per hour. The
144 average duration of a complete tour of the 315 m² exposition
145 area was planned for 15 min.

After agreeing on one of the official languages of *Expo.02* 146
(English, French, Italian, or German), the robot started moving 147
to the exhibits like Industry robot (A), Medical robot (B), Fossil 148
(D) (showing body implants), or mechanical underwater toys at 149
Aquaroids (E). Visitors could control the miniature robot Alice 150
(F) using buttons on the tour-guide robots. Other exhibits like 151
Face Tracking (K) and our Supervision Lab (M) or the robot 152
presentation of itself Me, myself and I (C) gave some insight to 153
the mobile robots' perception of the environment. 154

The tours were dynamic, in that the exhibits presented were 155
chosen by the visitor. After completing the presentation of one 156
exhibit, robots requested a list of free exhibits. To promote 157
visitor flow toward the exit, only free exhibits, located closer 158
to the exit than the current could be selected by the visitors. 159
A tour ended after a fixed number of exhibits, with the robot 160
saying goodbye and returning to the welcome area. 161

Some robots were dedicated to one exhibit and interacted 162
without the need to give a tour: the Presenter robot (G), 163
explaining the inner workings of a robot, the Jukebot (H), 164
proposing a selection of music, the Philosopher (J), speaking 165
about good and the world, and the Photographer (L), taking 166
pictures and displaying them on three television towers, the so- 167
called Cadavre Exquis (N). 168

169 II. TOUR GUIDE: ROBOX

The autonomous mobile system RoboX was developed for 170
Expo.02 at the Autonomous Systems Lab and produced by its 171
spin-off company BlueBotics SA. It is shown in Fig. 2. Safe 172
and reliable operation was mandatory for its use in a public 173
exposition, in close proximity to hundreds of visitors. For most 174
of the visitors, RoboX was the first contact with a real robot. 175
This called for friendly appearance and an intuitive operation. 176
How visitors would react toward an autonomous machine was 177
difficult to predict. Thus, considerable effort was undertaken to 178
make the robot robust against destructive behavior. 179

180 A. *Hardware*

In order to ensure that visitors could easily spot RoboX 181
even in crowded settings, the robot's height is 1.65 m. Heavy 182
components are in its mobile base, which has a diameter of 183
0.70 m (0.90 m with foam bumpers), giving the robot good 184
equilibrium. The battery pack provides up to 12 h of autonomy 185
and makes up a large part of the system's weight of 115 kg. 186
RoboX has two differentially driven wheels on its middle axis, 187
which allows turning on the spot. This is a key feature when 188
visitors are blocking its way. 189

The mobile base contains the following: two laser range 190
finders (Sick LMS 200); the drive motors; the safety circuit; and 191
the tactile bumpers. Additionally, the two computers making 192
the robot autonomous, a PowerPC 375 MHz running XO/2 and 193
a personal computer (PC) Pentium III 700 MHz running Win- 194
dows 2000, are located there. To interact with visitors, RoboX 195
provides a mechanical face with a Firewire color camera and 196
a light-emitting diode (LED) matrix, two loudspeakers, and 197
interactive buttons. Two robots were equipped with a direc- 198
tional microphone matrix (Andrea Electronics DA-400 2.0) for 199

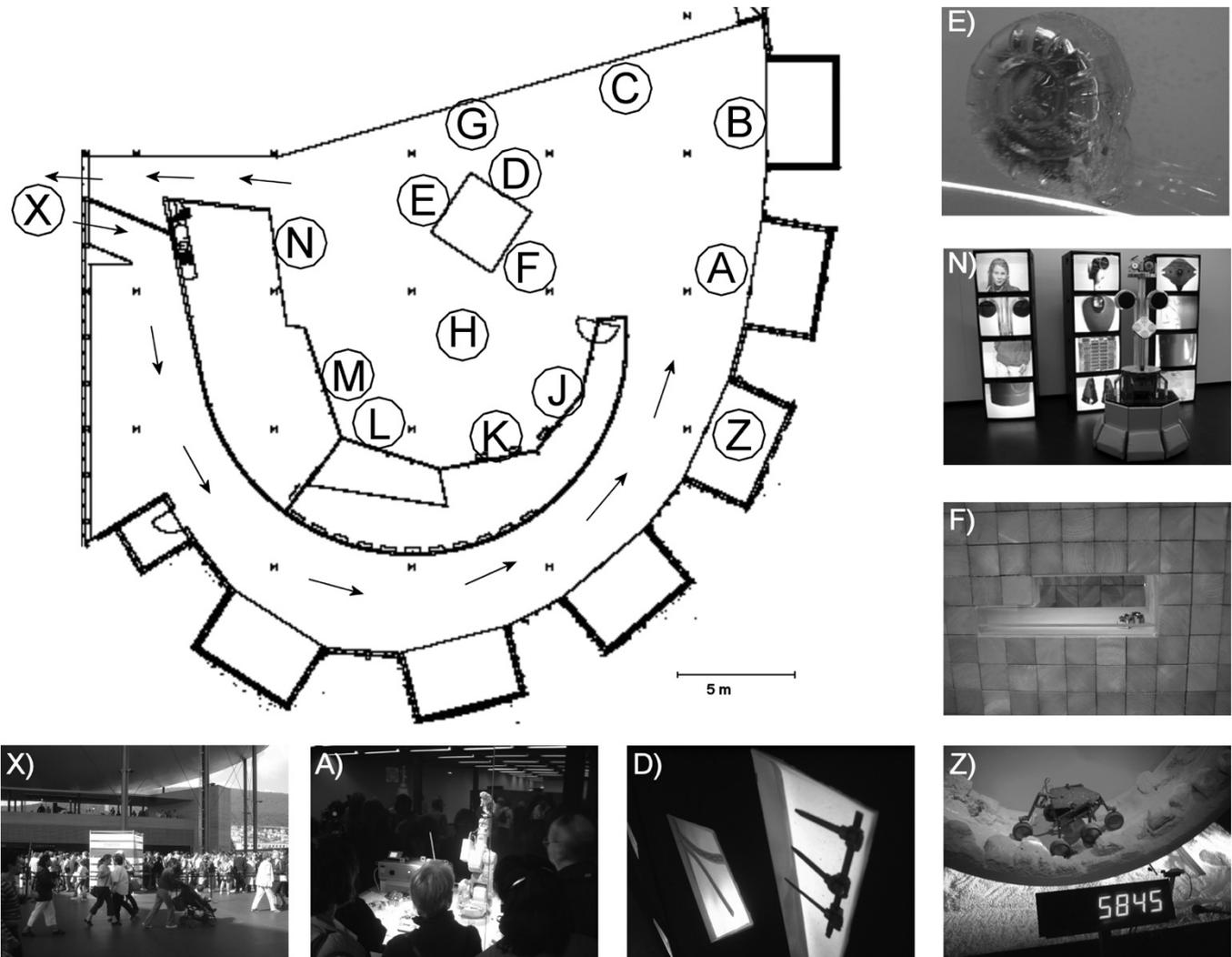


Fig. 1. Overview of the *Robotics* exhibition at *Expo.02*. The plan in the upper left indicates the location of exhibits and other places of interest. The insets are labeled accordingly, as well as some references in the main text. Exhibits A–N were parts of guided tours (exhibit Z was added to this 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.

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

202 B. Navigation

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

213 Localization is based on line features extracted from
214 laser range data, with multiple hypotheses tracked using a
215 Kalman filter [15]. It was designed for operation in unmod-

ified environments and performs well in cluttered situations. 216
Using line features keeps the map compact and computational 217
costs low. 218

Motion control combines several approaches, in a manner 219
similar to the following [16]: NF1 [17] for local path planning; 220
elastic bands [18] as adaptive path representation; and the 221
dynamic window approach [19] for obstacle avoidance. The 222
method has high computational efficiency due to lookup tables 223
similar to [20]. More details can be found in [21]. 224

C. Safety

225

Robot components that influence motion are defined as 226
safety critical, namely: speed control; obstacle avoidance; 227
laser scanner; and bumpers. All those are running on the 228
RTOS of the PowerPC. Taking into account the possibility 229
of a failure of the PowerPC, a redundant safety controller is 230
added. It is implemented using a peripheral interface controller 231

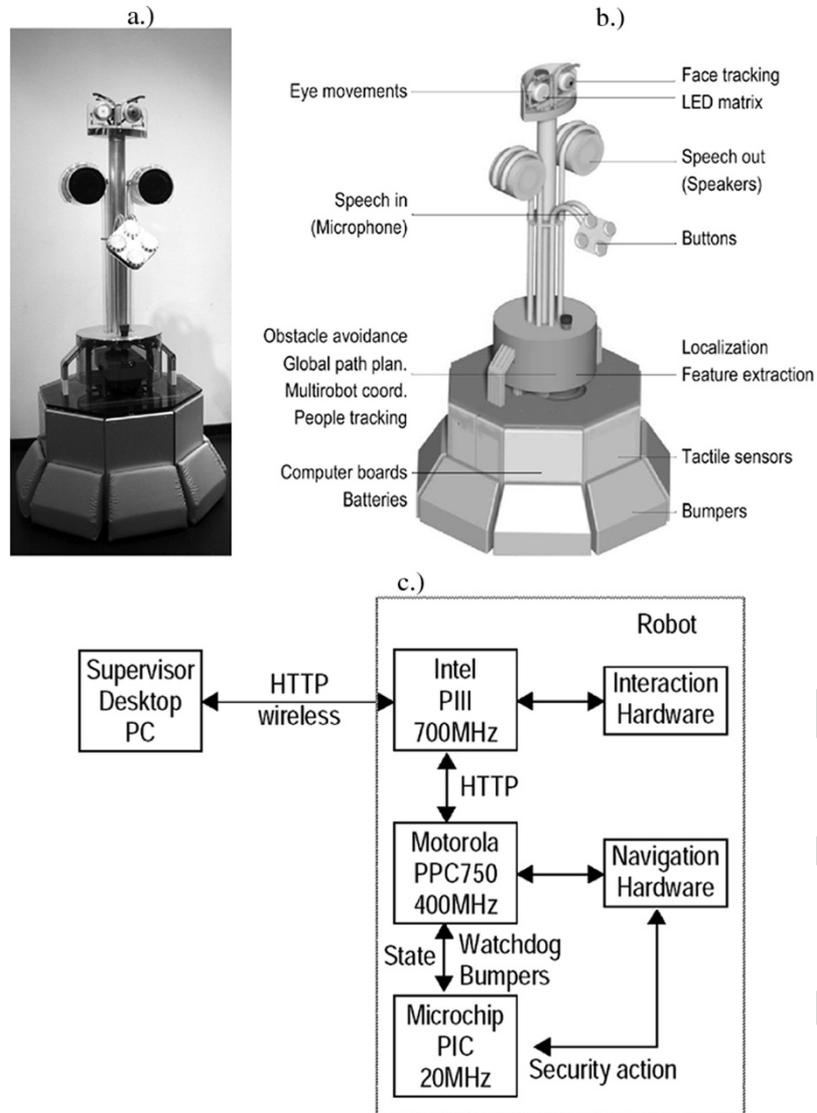


Fig. 2. (a) Interactive mobile robot RoboX. (b) Navigation and interaction elements of RoboX. (c) RoboX safety system layout: Navigational components on the RTOS of the PowerPC, Windows 2000 contains interactive components only (i.e., not safety critical). The PIC microcontroller serves as a watchdog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

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

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

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

III. MODALITIES FOR INTERACTION

249

In an exhibition, the tour-guide robot interacts with individ- 250
 251 ual visitors as well as crowds of people. In both situations, it
 is important that RoboX takes the initiative. Thus, a primary 252
 component of a successful tour guide is the ability to engage 253
 in a meaningful conversation in an appealing way [23]. High- 254
 performance environmental perception and intuitive expressive 255
 elements are the means used to achieve this goal. 256

In the following, the modalities for interaction are presented 257
 and their main features described. We distinguish perceptive 258
 and expressive modalities. 259

A. Perceptive Modalities

260

RoboX is equipped with multiple sensors. A camera and two 261
 laser scanners give the robot a sense of people surrounding it, an 262
 important skill for interaction as reported in other public space 263

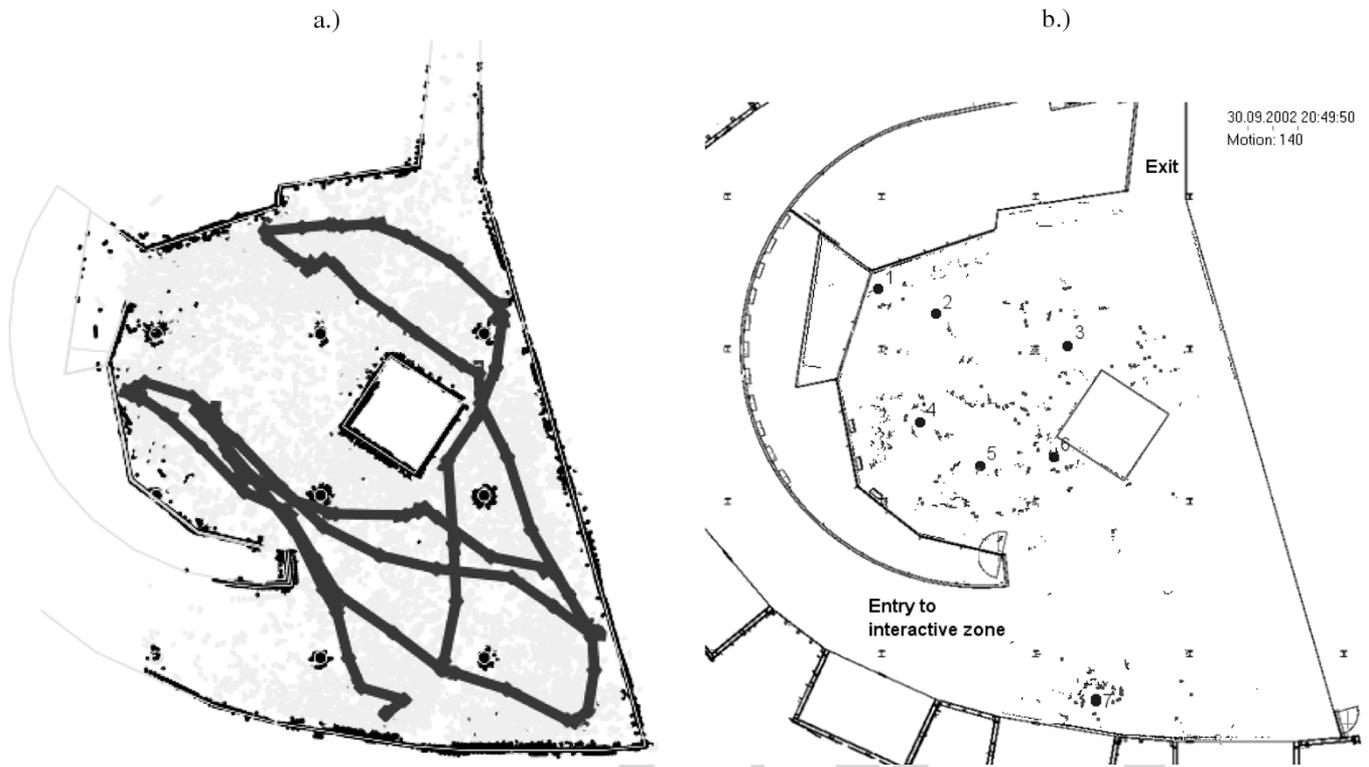


Fig. 3. Motion detection using laser range finder data from a mobile platform at *Expo.02* while roaming the 315 m² exhibition area. (a) The path of the robot during 17 min with light points indicating dynamic parts and dark points representing static parts. (b) Snapshot of the exposition with data from several robots. One hundred forty motion elements are detected at this moment.

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

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

273 1) *Motion Detection*: Motion is detected in order to find
 274 people in the robot’s vicinity. Other methods could be em-
 275 ployed, e.g., using shape information [25], [26] or singularities
 276 in the environment [27]. Our method is presented in detail in
 277 [28]–[30].

278 A result of the algorithm is shown in Fig. 3(a). The envi-
 279 ronment is assumed to be convex and static in the beginning.
 280 The range readings are integrated into the so-called static map,
 281 consisting of all currently visible elements that do not move.
 282 Only one information is stored for each angle. In the next step,
 283 the new information from the range finder is compared with
 284 the static map. Assuming a Gaussian distribution of the sensor
 285 readings representing a given element, a chi-square test can be
 286 used to decide whether the current reading belongs to one of
 287 the elements of the static map or originates from a dynamic
 288 object. All static readings are used to update the static map.
 289 Readings labeled as dynamic are used to verify the static map
 290 as follows: If the reading labeled as dynamic is closer to the
 291 robot than the corresponding value from the static map, the
 292 latter persists. In case it is farther away than the map value,

it is used to update the map, but remains labeled as dynamic. 293
 All dynamic elements are clustered according to their spatial 294
 location. Each cluster is assigned a unique identification (ID) 295
 and the center of gravity of its constituting points in Cartesian 296
 space is computed. The classification, update, and validation 297
 steps are repeated for every new scan. In case of robot motion, 298
 the static map is warped to the new position. 299

2) *Face Tracking*: Fig. 4 shows an example of face tracking 300
 based on red green blue (RGB) data of the camera located in 301
 the robot’s left eye. Skin-colored regions are extracted using an 302
 algorithm presented in [31] and [32]. To reduce the sensitivity 303
 against illumination, green and blue are normalized using the 304
 red channel. Then, fixed ranges for blue, green, and brightness 305
 are accepted as skin color. Taking brightness into account 306
 rejects regions of insufficient saturation. Erosion and dilation 307
 remove small regions from the resulting binary image. The 308
 binary image is clustered and the contour of each cluster is 309
 extracted. Heuristic filters are applied to suppress skin color 310
 regions that are not faces. These filters are based on rectangular 311
 areas, their aspect ratio, and the percentage of skin color 312
 within the rectangle. Clusters are linked over time using the 313
 nearest-neighbor assignment. Clusters that remain unassigned 314
 to previous tracks are added and tracked until they leave the 315
 camera’s field of view. 316

Information gathered from the face tracker is used in several 317
 interaction parts. Together with motion tracking, it helps to 318
 verify the presence of visitors and to orient the robot’s face 319
 toward the user. Furthermore, it triggers the behavior engine 320
 emotional state machine, which is presented in Section IV. 321



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.

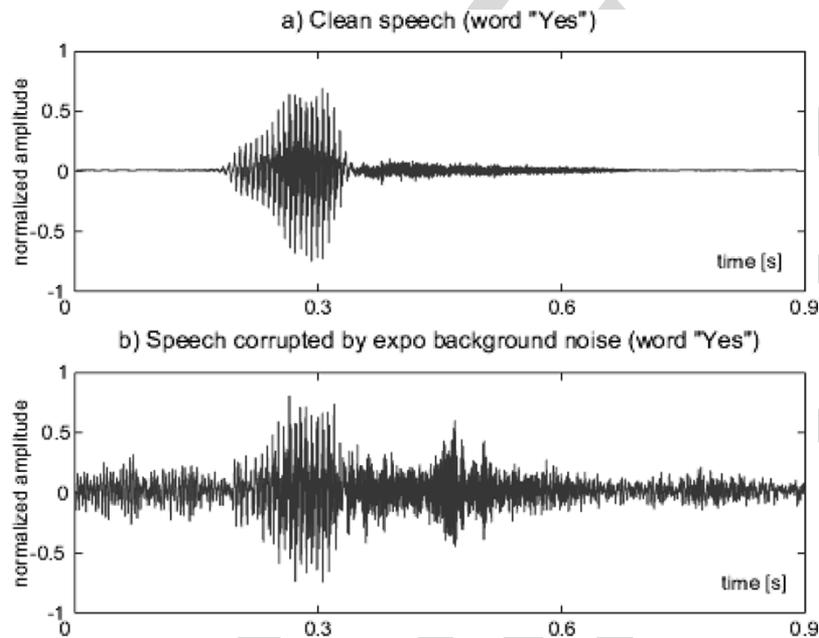


Fig. 5. Samples of the word Yes under (a) quiet and (b) noisy conditions of the exhibition room.

322 3) *Speech Recognition*: A primary requirement of *Expo.02*
 323 was that the tour-guide robots should be capable to interact
 324 with visitors using four languages: French, German, Italian,
 325 and English. The large number of visitors prohibited the use
 326 of handheld microphones as in [10], the adopted solution was
 327 to mount a microphone array on the robot.
 328 Studying related work on tour-guide robots led us to the
 329 following observations [33]. First, even without voice-enabled
 330 interfaces, tour-guide robots are very complex, involving sev-
 331 eral subsystems that need to communicate efficiently in real
 332 time. This calls for speech interaction techniques that are easy
 333 to specify and maintain, and that lead to robust and fast speech
 334 processing. Second, the tasks that most tour-guide robots are
 335 expected to perform typically require only a limited amount
 336 of information from the visitors [34]. These points argue in
 337 favor of a very limited but meaningful speech recognition
 338 vocabulary and for a simple dialogue management approach.

The solution adopted is based on yes/no questions initiated by
 339 the robot where visitors' responses can be in the four required
 340 languages (oui/non, ja/nein, si/no, yes/no). This simplifies the
 341 voice-enabled interface by eliminating the specific speech un-
 342 derstanding module and allows only eight words as multilingual
 343 universal commands. The meaning of these commands depends
 344 on the context of the questions asked by the robot. A third
 345 observation is that tour-guide robots have to operate in very
 346 noisy environments, where they need to interact with many
 347 casual persons (visitors). Fig. 5 presents typical speech samples
 348 from quiet and noisy conditions. In the exhibition room, the
 349 signal is drowned in babble combined with the noise of robot
 350 movement and beep sounds. This calls for speaker-independent
 351 speech recognition and for robustness against noise. The first
 352 task of the speech recognition event is the acquisition of the
 353 useful part of the speech signal. The adoption of acquisition
 354 limited in time (3 s) is motivated by the average length of yes/no
 355

356 answers. Ambient noise in the exhibition room is among the
 357 main reasons for speech recognition performance degradation.
 358 A microphone array (Andrea Electronics DA-400 2.0) is used
 359 to add robustness without additional computational overhead.
 360 During the 3-s acquisition time, the original acoustic signal
 361 is processed by the microphone array. The mobility of the
 362 tour-guide robot is very useful for this task since the robot,
 363 when using the motion detection system, can position its front
 364 in the direction of the closest visitor and, thus, directs the
 365 microphone array. The preprocessing of signals of the array
 366 includes spatial filtering, dereverberation, and noise canceling.
 367 This preprocessing does not eliminate all the noise and out-
 368 of-vocabulary (other than yes/no) words. It provides sufficient
 369 quality and nonexcessive quantity of data for further process-
 370 ing. Recognition should perform equally well on native and
 371 foreign speakers of the target language. We are interested in
 372 a low error rate and rejection of irrelevant words. At the heart
 373 of the robot's speech recognition system lies a set of algorithms
 374 for training statistical models of words subsequently used for
 375 the recognition task. The signal from the microphone array is
 376 processed using a Continuous Density Hidden Markov Model
 377 (CDHMM) technique where feature extraction and recognition
 378 using the Viterbi algorithm are adapted to a real-time execution.
 379 It offers the potential to build word models for any speaker
 380 using one of the mentioned languages and for any vocabulary
 381 from a single set of trained phonetic subword units. The major
 382 problem of a phonetic-based approach is the need for a large
 383 database required for training a set of speaker-independent
 384 and vocabulary-independent phoneme models. This problem
 385 was solved using standard European and American databases
 386 available from our speech processing laboratory, as well as
 387 specific databases with the eight keywords recorded during
 388 experiments. Four language-specific databases were used to
 389 train four sets of phoneme-based subword models. Training
 390 employed the CDHMM toolkit HTK [35] based on the Baum-
 391 Welch algorithm. Out-of-vocabulary words and spontaneous
 392 speech phenomena like breath, coughs, and all other sounds that
 393 could cause a wrong interpretation of visitor's input also have
 394 to be detected and excluded. For this reason, a word spotting
 395 algorithm with garbage models has been added to the recogni-
 396 tion system. These garbage models were built from the same set
 397 of phoneme-based subword models [36], [37], thus, avoiding
 398 an additional training phase or software modification. Finally,
 399 the basic version of the system was capable of recognizing
 400 yes/no words in the required languages and acoustic segments
 401 (undefined speech input) associated with the garbage models.

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

409 B. Expressive Modalities

410 When RoboX finds people in close distance, it should greet
 411 and inform them of its intentions and goals. The most natural



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

and appealing way to do this is by speaking. In addition to
 speech, a large number of facial expressions and body move-
 ments are used in human communication to enhance the mean-
 ing of the spoken dialogue. Additional expression is conveyed
 by varying prosodic parameters.

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 ad-
 ditional 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 prob-
 lems 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).

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.

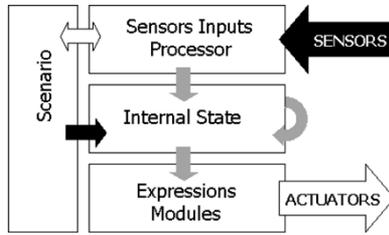


Fig. 7. Information flow: The scenario program is executed and influenced by sensor input. The internal emotional state is influenced by signals from several sources, including the scenario. RoboX expression results as a function of its internal state.

460 The use of prerecorded samples was ruled out by the require-
 461 ment of conveying the robot's emotional state by modulating
 462 speech parameters, and to allow dynamic generation of spoken
 463 sequences. RoboX employs speech synthesis system based on
 464 LAIPTTS [39], [40] and Mbrola [41] for French and German,
 465 whereas English and Italian were synthesized using ViaVoice
 466 [42]. Prosodic parameters as pitch, volume, and rate can be
 467 changed while the robot is speaking.

468 IV. EMOTIONAL STATE MACHINE

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

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

481 A. Template Expressions

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

487 For each template expression, a parameter set for the expres-
 488 sive modalities was defined manually. Table I shows the para-
 489 meter sets qualitatively. We chose to mimic human expressions
 490 and to exaggerate them where possible, given the capacities of
 491 the robot.

492 To create a more lively appearance, these template expres-
 493 sions allow the definition of a value range for the expressive
 494 parameters. Within this range, the actual output is defined ran-
 495 domly and changes continuously. The emotional state machine
 496 provides the scenario with a control on how these parameter
 497 ranges are used:

498 1) Default behavior: Only eyebrows are controlled by the
 499 emotional state machine. Their position is changed ac-
 500 cording 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. 501
 Those affect the gaze direction and speed of movement in 502
 function of the robot's mood. The gaze direction tells a lot 503
 about the state of mind of human beings. We, therefore, 504
 determine a specific window for the random movement in 505
 the eye space, which is shown in Fig. 8. 506
- 3) Random sequences: For each template expression, a set 507
 of movements using eyebrows and eyes can be imple- 508
 mented, e.g., the LED matrix may show a teardrop among 509
 other symbols when the robot is sad. 510

B. Mapping Perception to Affects 511

The sources taken into account in creating expressions com- 512
 prise of the following: face tracking; motion detection; buttons; 513
 laser scanners; bumpers; and battery. For different conditions, 514
 these sources are evaluated with respect to the goals of the 515
 robot. The resulting mapping of conditions to desired expres- 516
 sions is shown in Table II. In order to display these expressions, 517
 the source information is used to change the internal emotional 518
 state, ensuring a smooth transition. 519

If the robot cannot fulfill its task, it becomes unhappy 520
 (sorrowful when nobody is in sight during a presentation; angry 521
 if someone plays with the buttons disturbing the robot, or 522
 when someone completely blocks the way). The robot is happy 523
 when successfully making its job (joyful when seeing someone 524
 during a presentation). 525

C. Representation of the Emotional State 526

When inputs require the emotional state to change, the 527
 expression changes accordingly. It is not credible for all ex- 528
 pressions to change instantaneously from, e.g., happy to sad. 529
 To do so, we derive a set of intermediate expressions as an in- 530
 terpolation of template expressions, where the transition speed 531
 depends on the new emotional state. 532

We use the three-dimensional (3-D) Arousal-Valence- 533
 Stance (AVS) space [44] as an internal representation of the 534
 emotional state (see Fig. 9). The advantage of AVS space is that 535
 it can be easily mapped to the expression space for the seven 536
 template expressions. 537

Transition in this space results from signals from several 538
 sources or explicit scenario inputs, which are transformed to a 539
 point of the AVS space \vec{a}_{input} . The new affect \vec{a}_{new} is computed 540

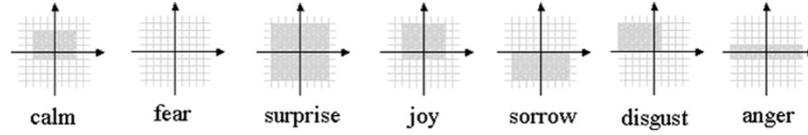


Fig. 8. Parameter range of eye position (pan, tilt) for different template expressions.

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

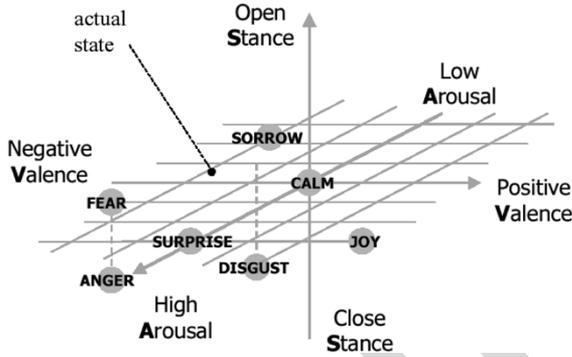


Fig. 9. The robot's emotional state is a point in the AVS space. The robot's seven template expressions are specific states in this space, corresponding to specific output parameters on the expressive modalities. Transitions from one state to another pass through nonmodeled intermediate expressions,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

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

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

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

548 D. Expression Generation

549 The parameter set \vec{p}_{new} for the new expression, which
550 is displayed, is a weighted mean of the parameter sets \vec{p}_e
551 for the seven template expressions, denoted as E . The inverse

of the distance of the current state \vec{a}_{new} to the template states
 \vec{a}_e is the weight w_e . The new parameter set is given by (2)

$$w_e = (1 + \|\vec{a}_{new} - \vec{a}_e\|)^{-1}$$

$$\vec{p}_{new} = \frac{1}{\sum_E w_e} \sum_E w_e \vec{p}_e. \quad (2)$$

Intuitively, the closer the current state is to the center of a 554
template expression, the more the current expression reflects
555 that emotional state. Transitions from one expression to another
556 do not need to be modeled explicitly, but result from the state
557 transition in the affect space as shown in Fig. 10. 558

559 V. INTERACTIVE SCENARIOS

Interactive scenarios are the combination of stage-play pre- 560
sentations and reactive scenarios. By reactive scenarios, we 561
mean small dedicated programs for special situations. Fig. 7 562
gives an overview of the interactive system. 563

The scenario composition explains how to create stage-play 564
scenarios for presenting exhibits and reactive scenarios for 565
special situations (robot blocked, battery low). The scenarios 566
may influence the expression directly, by requesting a certain 567
emotional state, or rely on a continuous interpretation of the 568
sensor data to generate expressions. 569

Stage-play scenarios can combine modalities for interaction 570
(Fig. 11) to create presentations [Fig. 12(a)]. 571

In their simplest form, stage-play scenarios are a linear suc- 572
cession of commands. Introducing parallel execution of tasks 573
increases the scenario's complexity, for instance, allowing to 574
change the facial expression while speaking. Even more com- 575
plex scenarios contain branches. Such decisions may depend 576
on speech recognition [see the example in Fig. 12(a)], motion 577
detection, or button events. 578

Two kinds of scenarios are used, namely: 1) presentation 579
scenarios; and 2) reactive scenarios. Depending on the inter- 580
action strategy, presentation scenarios are used as a set to create 581
a tour, or dedicated for one application. Presentation scenarios 582
in a tour are executed depending on visitor choices and the 583
availability of free exhibits. 584

The emotional state machine may inject reactive scenarios 585
into the program, if required, even when a presentation scenario 586
is already running. 587

When a reactive scenario is triggered, the main program 588
dynamically changes the current presentation scenario. The 589
corresponding reactive scenario is executed until the robot can 590
continue the tour. It is possible to load a number of different 591
scenarios for each case, which allows the robot to vary com- 592
ments, if the situation did not change after execution of the first 593
reactive scenario. 594

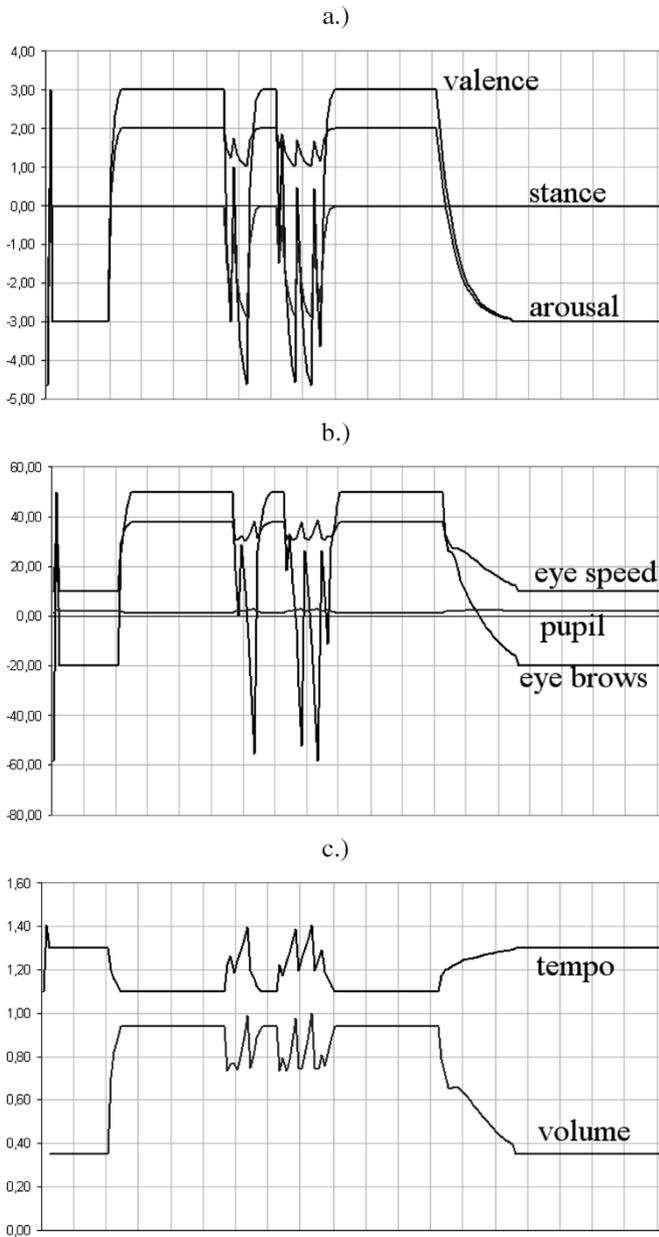


Fig. 10. Relation between affect and expressive modalities during a short experiment. (a) Affect change in the AVS space over time. (b) Parameters for eyes in percent of their maximal value over time. (c) Parameters for synthesized speech, where 1.0 is the default value for volume and speed. In the beginning, 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.

595 A. Presentation Scenario

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

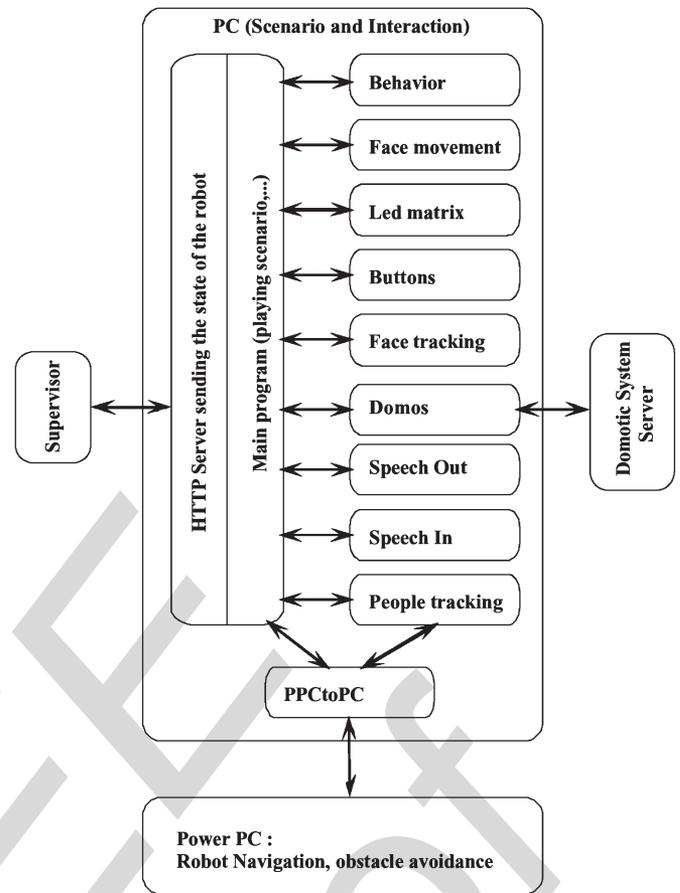


Fig. 11. Block diagram of the main modalities for interaction and how they are linked. Three interfaces function as gateways, namely: 1) the supervision computer; 2) the control of the environment through a dedicated server (Domos); and 3) the navigation part of the robot.

B. Reactive Scenario

603

The reaction of RoboX to different situations is programmed
604 with respect to the goals and needs of the tour. For example,
605 if a visitor is blocking the path, RoboX shows anger, because
606 this delays the tour. Cases for which reactive scenarios were
607 developed are as follows: batteries are running low; someone is
608 playing with the buttons; the robot is blocked; and the bumpers
609 are touched. An example is given in Fig. 12(b). It is started
610 when the robot is blocked.
611

VI. RESULTS

612

The exposition *Expo.02* took place from May until October
613 2002. *Robotics* was one exhibition among several related to
614 different topics. It was open to the public 10 h a day and 12 h
615 during the last month.
616

The visitors typically spent 10–30 min in the *Robotics@*
617 *Expo.02* exhibition. This classifies the man–machine contact
618 as short-term interaction, where the visitors, in contrast to
619 the exposition staff, did not have enough time to form a
620 deeper relationship with the robots as in the experiments re-
621 ported in [13].
622

We will report on the overall performance of the robots
623 during the exposition. We try to assess the quality of the
624

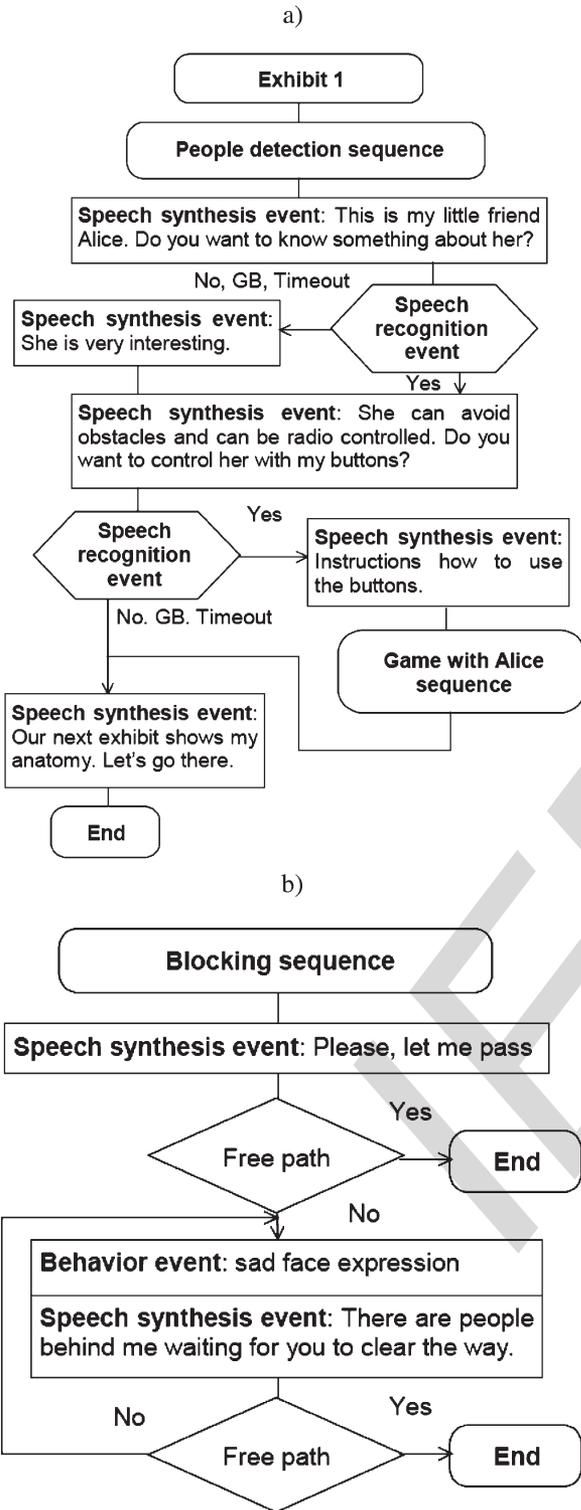


Fig. 12. (a) Sequence presenting the exhibit Alice using people detection, speech synthesis, and recognition. (b) Reactive scenario, which is used when the robot is blocked. When visitors keep RoboX from reaching a goal, it changes its expression. If the obstruction persists, RoboX complains until the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

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

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

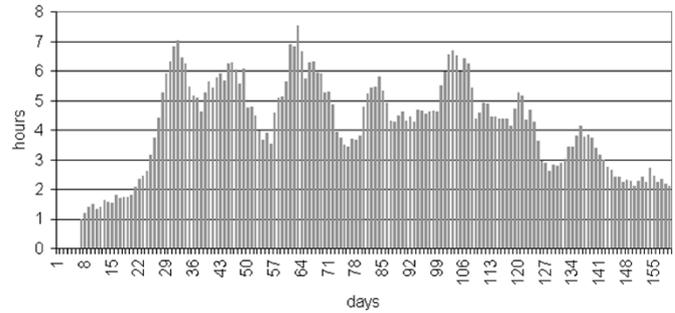


Fig. 13. MTBF as average of 11 robots for each day of the exposition. Note the improvement of MTBF during the first 30 days from 1 to 7 h. During the last month of the exposition, the MTBF drops again. At the same time, the opening time of the exposition was raised from 10 to 12 h, increasing wear on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

report on observations made in the exposition related to these 629
630 modifications.

A. Robot Performance During the Exposition

631
632 During *Expo.02*, 11 RoboXs were guiding more than 686 000
633 visitors through *Robotics*. Everyday, between 6 and 11 robots
634 were running a 10-h shift each. On the average, 8.4 robots
635 were interacting with 4317 visitors per day (minimum = 2299
636 and maximum = 5473 visitors), adding up to the following
637 operational values:

- 1) total run time: 13 313 h; 638
- 2) total motion time: 9415 h; 639
- 3) traveled distance: 3316 km; 640
- 4) maximum speed: 0.6 m/s; 641
- 5) average speed: 0.098 m/s; 642
- 6) average interactions: 51 visitors/robot/h; 643
- 7) mean time between failure (MTBF): 3.26 h. 644

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

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

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

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

666 Regarding the safety aspects, we neither received complaints
667 nor did we observe any dangerous situations. Accidents did not
668 occur. When not obstructed intentionally by visitors, obstacle

669 avoidance was able to guide RoboX, even in tight situations
 670 without collision. Of course, intentional obstructions occurred.
 671 The low speed of RoboX and its immediate stopping on contact
 672 made blocking the robot's way a popular and harmless game
 673 for visitors.

674 B. Results From Survey

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

- 678 1) How do you rate the robot's appearance?
- 679 2) How do you rate the robot's character?
- 680 3) How good is the synthesized speech?
- 681 4) How did you learn to use the robot?
- 682 5) How do you rate the speech recognition? (only on two
 683 robots)
- 684 6) Which sensor is used for navigation?
- 685 7) Which exhibits did you visit?
- 686 8) How do you rate the exhibition?
- 687 9) Would you prefer a normal information desk or an inter-
 688 active robot when asking for directions?

689 Answers were collected from 209 visitors, 106 (58%) female
 690 and 89 (42%) male, speaking German 128 (61%), French 75
 691 (36%), or Italian 6 (3%). The average age was 34.4 years, the
 692 oldest participant was 74 years old, and the youngest was five
 693 years old.

694 The aggregated results to questions 1, 2, and 8 show a
 695 very similar distribution as follows: very good (20%); good
 696 (51%); acceptable (26%); bad (3%) within a small margin (3%).
 697 This strongly suggests that, during the short time of their stay,
 698 visitors perceived the robots, probably the entire exposition as
 699 a whole.

700 Speech synthesis (question 3) was rated above the overall
 701 average with a distribution as follows: very good (31%); good
 702 (44%); satisfactory (24%); and bad (1%). The same applies
 703 for speech recognition (question 5) with a distribution as
 704 follows: very good (37%); good (39%); satisfactory (20%);
 705 and bad (4%).

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

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

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

719 C. Evaluation of Modalities for Interaction

720 Regarding the modalities for interaction, we were inter-
 721 ested in the reliability of motion detection, face tracking, and
 722 speech recognition under *Expo.02* conditions. Concerning the

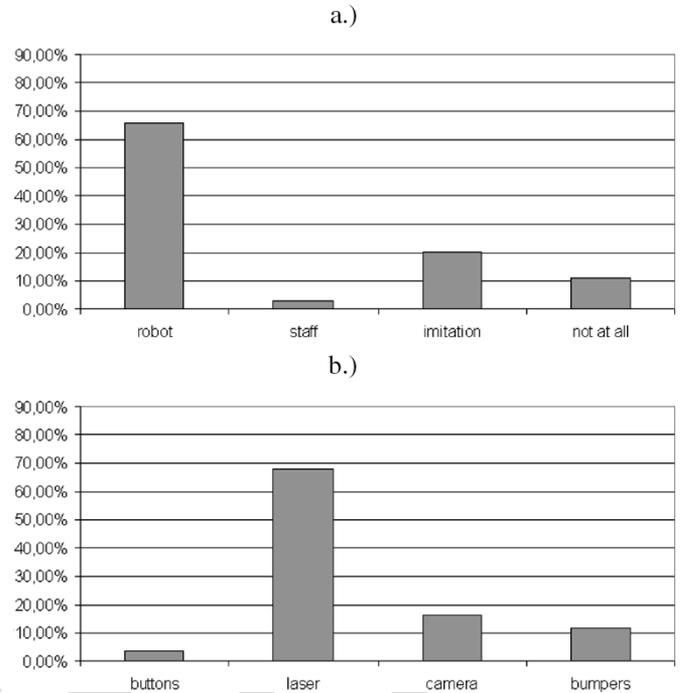


Fig. 14. Results from the survey. Only one selection was possible. (a) How did the visitors learn how to use the robot? The answers from the visitors show that the robot itself was the best teacher. Note that only 11% of the visitors did not learn how to use the robot. (b) Understanding of elementary 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%

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

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

755 2) *Face Tracking*: The performance of the face tracking
756 algorithm was evaluated quantitatively from a sequence of im-
757 ages, similar to the one shown in Table IV. The sequence lasting
758 11 min was sampled at 4 Hz resulting in 2800 images. The
759 manual evaluation of the faces present, detected and tracked per
760 image, was limited to every twentieth image, since consecutive
761 images are very similar. In total, 169 images were classified.
762 The results are summarized in Table IV. Images were classified
763 in categories. We distinguish images as follows: sharp images;
764 images with motion blur; and dark images. The dark image
765 class comprises a part at the beginning of the sequence with
766 very low illumination, for which the skin color model was not
767 designed.

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

772 In the 169 images evaluated, a total of 1047 faces were
773 present, of which 497 were correctly detected. A total of
774 37 regions were detected, which did not correspond to a face,
775 resulting in a type-II error of 6.9%. The detection rate was
776 47.5% on the average and 64.2% for sharp images. The detec-
777 tion rate drops to 12.59% for dark images. This is probably
778 due to the skin color model, which was created for normal
779 illumination.

780 For motion detection, the type-I error increases again with
781 the number of persons present, probably due to partial occlu-
782 sions. The detection rate of 47.5% (64.2% sharp images) is
783 in part due to the crowded situation of up to 11 faces on the
784 images, which cover a considerable smaller angle than the laser
785 sensors. The type-II error is still low (8.9%), so that RoboX

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%

almost never assumed the presence of a person, when, in fact, there was none.

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 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 were 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

AQ3

AQ4



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%

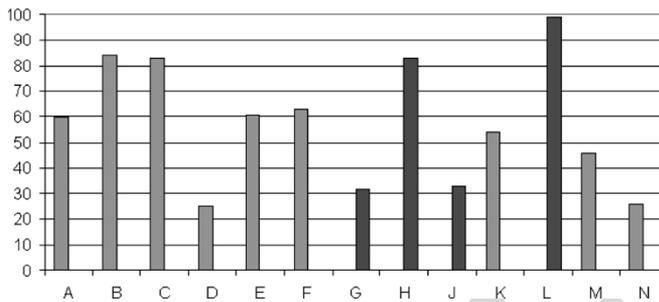


Fig. 16. Number of visitors per exhibit. Exhibits are arranged according to their distance from the entry. Dark bars indicated the robots as exhibits and lighter bars indicate the tour-guide exhibits. The corresponding locations are shown in Fig. 1. There are strong variations between both groups. It 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.

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

834

VII. DISCUSSION

835 The discussion comprises an assessment of interaction strat-
836 egy by means of visitor density, a report on the evolution of
837 scenarios and changes in the exhibition, and personal impres-
838 sions from staff members, who worked in *Robotics@Expo.02*
839 throughout the 5-month period.

A. Interaction Strategies and Visitor Density

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

As was pointed out earlier, visitors perceived the exhibition 845
as a whole, making it difficult to evaluate different types of 846
interaction directly with a survey. However, visitors correctly 847
remembered which part of the exposition they visited. We argue 848
that the number of visitors per exhibit indicates its popularity 849
and try to infer from this which types of interactions were 850
appealing to visitors. 851

Particular interest received: Photobot (L) and Jukebot (J), 852
which were not part of the guided tour, but were served by a 853
dedicated RoboX. Among the tour stations, two of the three 854
foremost stations received the most attention [Medical robot (B) 855
and Me, myself and I (C)]. 856

Visitors started the exhibition by joining a guided tour pro- 857
vided by the robots. With the exception of Fossil (D), the 858
number of persons per guided group decreased gradually 859
toward the exit, probably because they were attracted to other 860
parts of the exhibition. Our observations throughout *Expo.02* 861
confirm the visitor distribution derived from the survey and 862
shown in Fig. 16. In our opinion, the lack of visitors at Industrial 863
robot (A) was due to its proximity to the welcome area. Visitors 864
sometimes started tours inadvertently, selecting the wrong lan- 865
guage. Instead of following the robot, they joined another tour 866
in their language given by one of the other robots nearby. In 867
fact, when we moved the welcome area from around point (A) 868
into the hallway near point (Z), more visitors were attracted to 869
Industrial robot. 870

The Fossil (D) exhibit was presented using the same tech- 871
niques as Medical robot (B), Me, myself and I (C), and 872
Auaroids (E). The lack in visitors may be attributed to its 873
location as it is not in the exhibition's mainstream. This may 874
as well apply to the Presenter robot (G) located nearby, which 875
was explaining some insights of RoboX using projected slides. 876
Stations that explained robot perception were Face Tracking 877
(K) and Supervision Lab (M). 878

The noticeable interest in the exhibits Photobot (L) and 879
Jukebot (H) convinced us that short and highly reactive scenar- 880
ios create an interesting interaction for the visitor, since their 881
actions were immediately rewarded by the robot. 882

B. Scenario Evolution

883
884 Stage-play scenarios were revised throughout *Expo.02*, re-
885 flecting experience gathered during the exhibition. As an exam-
886 ple of this evolution, the introduction scenario is outlined. Then

887 we address the issue of timing with regards to visitor behavior
888 and robot reaction.

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

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

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

908 Similar problems arose for introduction scenario using but-
909 tons. It started with the question sequence “red—French/
910 blue—German/green—English/orange—Italian”. When saying
911 “red for French,” some visitors immediately pressed on the red
912 alarm button instead of waiting for the end of the sentence and
913 choosing by pressing on the red colored button.

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

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

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

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

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

941 Timing was also found to be an issue when combining stage-
942 play and reactive scenarios. Sometimes, events like touching
943 the buttons occurred while the robot was in the middle of a
944 long task; when it finally responded to the event after task



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.

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.

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).

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 opera-
tional performance.

AQ5

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

981 VIII. CONCLUSION

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

988 The objectives of interaction, the exhibition, and its develop-
 989 ment have been presented. Robotic modalities for interaction
 990 have been presented in detail. Perceptive elements (motion
 991 detection, face tracking, speech recognition, buttons) have been
 992 distinguished from expressive ones (robotic face, speech syn-
 993 thesis, colored button lights). An approach for combining stage-
 994 play and reactive scenarios has been presented. An emotional
 995 state machine has been used to create convincing expressions
 996 from the robot.

997 For the entire 5-month duration of the exhibition, an evalu-
 998 ation of the robot performance has been given. A performance
 999 analysis of modalities for interaction has also been presented.
 1000 Survey results to assess human-robot interaction and interac-
 1001 tion strategies have also been included.

1002 The event *Robotics@Expo.02* has greatly contributed to our
 1003 experience in the field of large-scale human-robot interaction.
 1004 We hope that the results will contribute to the further develop-
 1005 ment of interactive robots.

1006 ACKNOWLEDGMENT

1007 The production of the 11 RoboXs was realized by Blue-
 1008 Botics, a spin-off of the Autonomous Systems Lab. The authors
 1009 thank all members of the team *Robotics@Expo.02* for their
 1010 outstanding contributions, namely: K O. Arras; M. de Battista;
 1011 S. Bouabdallah; D. Burnier; G. Froidevaux; X. Greppin;
 1012 B. Jensen; A. Lorotte; L. Mayor; M. Meisser; R. Philippsen;
 1013 P. Prodanov; R. Piguet; G. Ramel; M. Schild; R. Siegwart;
 1014 G. Terrien; and N. Tomatis. Apart from this core team, various
 1015 people from academia and industry supported the project. The
 1016 authors are particularly grateful to R. Philippsen for his help
 1017 in preparing the paper. The authors also thank P. Prodanov for
 1018 sharing his expertise on speech recognition and S. Vasudevan
 1019 for fruitful discussions.

1020 REFERENCES

- 1021 [1] M. Fujita, "AIBO: Toward the era of digital creatures," *Int. J. Rob. Res.*,
 1022 vol. 20, no. 10, pp. 781–794, Oct. 2001.
- 1023 [2] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and
 1024 K. Fujimura, "The intelligent ASIMO: System overview and integration,"
 1025 in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Lausanne,
 1026 Switzerland, 2002, vol. 3, pp. 2478–2483.
- [3] *The McGraw-Hill Illustrated Encyclopedia of Robotics and Artificial* 1027
Intelligence, McGraw-Hill, New York, Jan. 1995. 1028
- [4] C. Breazeal, "A motivational system for regulating human-robot inter- 1029
 action," in *Proc. Amer. Association Artificial Intelligence*, Madison, WI, 1030
 1998, pp. 54–61. 1031
- [5] B. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, 1032
 W. Steiner, and S. Thrun, "Experiences with an interactive museum tour- 1033
 guide robot," *Artif. Intell.*, vol. 114, no. 1–2, pp. 3–55, Oct. 1999. 1034
- [6] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, 1035
 D. Fox, D. Ahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, 1036
 "Probabilistic algorithms and the interactive museum tour-guide robot 1037
 Minerva," *Int. J. Rob. Res.*, vol. 19, no. 11, pp. 972–999, Nov. 2000. 1038
- [7] B. Graf, R. Schraft, and J. Neugebauer, "A mobile robot platform for as- 1039
 sistance and entertainment," in *Proc. 31st Int. Symp. Robotics*, Montreal, 1040
 Canada, 2000, pp. 252–253. 1041
- [8] I. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto, 1042
 "An effective mobile robot educator with a full-time job," *Artif. Intell.*, 1043
 vol. 114, no. 1–2, pp. 95–124, Oct. 1999. 1044
- [9] T. Willeke, C. Kunz, and I. Nourbakhsh, "The history of the Mobot 1045
 museum robot series: An evolutionary study," in *Proc. Florida* 1046
Artificial Intelligence Research Society (FLAIRS), Key West, FL, 2001, 1047
 pp. 514–518. 1048
- [10] R. Bischoff and V. Graefe, "Demonstrating the humanoid robot HERMES 1049
 at an exhibition: A long term dependability test," in *Workshop: Robots* 1050
Exhibitions, IEEE/RSJ Int. Conf. Intelligent Robots and Systems, Lau- 1051
 sanne, Switzerland, 2002. 1052 AQ6
- [11] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness 1053
 and attention in human-robot interaction," in *Proc. Amer. Association* 1054
Artificial Intelligence (AAAI) Fall Symp., Boston, MA, 2001. 1055 AQ7
- [12] —, "The role of expressiveness and attention in human-robot interac- 1056
 tion," in *Proc. IEEE Int. Conf. Robotics and Automation*, Washington, 1057
 DC, 2002, pp. 4138–4142. 1058
- [13] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "A practical experiment 1059
 with interactive humanoid robots in a human society," in *IEEE Int. Conf.* 1060
Humanoid Robots, Munich, Germany, 2003. 1061 AQ8
- [14] *Robotics@Expo.02*. [Online]. Available: <http://robotics.epfl.ch> 1062 AQ9
- [15] K. O. Arras, J. A. Castellanos, and R. Siegwart, "Feature-based multi- 1063
 hypothesis localization and tracking for mobile robots using geometric 1064
 constraints," in *IEEE Int. Conf. Robotics and Automation*, Washington, 1065
 DC, 2002, pp. 1371–1377. 1066
- [16] O. Brock and O. Khatib, "High-speed navigation using the global dynamic 1067
 window approach," in *Proc. IEEE Int. Conf. Robotics and Automation*, 1068
 Detroit, MI, 1999, pp. 341–346. 1069
- [17] J.-C. Latombe, *Robot Motion Planning*. Dordrecht, The Netherlands: 1070
 Kluwer, 1991. 1071
- [18] S. Quinlan and O. Khatib, "Elastic bands: Connecting path planning and 1072
 control," in *Proc. IEEE Int. Conf. Robotics and Automation*, Atlanta, GA, 1073
 1993, pp. 802–807. 1074
- [19] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to 1075
 collision avoidance," *IEEE Robot. Autom. Mag.*, vol. 4, no. 1, pp. 23–33, 1076
 Mar. 1997. 1077
- [20] C. Schlegel, "Fast local obstacle avoidance under kinematic and dynamic 1078
 constraints for a mobile robot," in *Proc. IEEE/RSJ Int. Conf. Intelligent* 1079
Robots and Systems, Victoria, Canada, 1998, pp. 594–599. 1080
- [21] R. Philippsen and R. Siegwart, "Smooth and efficient obstacle avoidance 1081
 for a tour guide robot," in *Proc. IEEE Int. Conf. Robotics and Automation*, 1082
 Taipei, Taiwan, 2003, pp. 446–451. 1083
- [22] R. Brega, N. Tomatis, K. Arras, and R. Siegwart, "The need for autonomy 1084
 and real-time in mobile robotics: A case study of xo/2 and pygmalion," 1085
 in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Takamatsu, Japan, 1086
 2000, pp. 1422–1427. 1087
- [23] P. Prodanov, A. Drygajlo, G. Ramel, M. Meisser, and R. Siegwart, "Voice 1088
 enabled interface for interactive tour guide robots," in *Proc. IEEE/RSJ* 1089
Int. Conf. Intelligent Robots and Systems, Lausanne, Switzerland, 2002, 1090
 pp. 1332–1337. 1091
- [24] S. Thrun, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, 1092
 D. Hähnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "MINERVA: 1093
 A second-generation museum tour-guide robot," in *Proc. Int. Conf.* 1094
Robotics and Automation (ICRA), Detroit, MI, 1999, vol. 3, pp. 1999– 1095
 2005. 1096
- [25] E. Prassler, J. Scholz, and E. Elfes, "Tracking people in a railway station 1097
 during rush-hour," in *Proc. Int. Conf. Computer Vision Systems (ICVS)*, 1098
 Las Palmas, Spain, 1999, pp. 162–179. 1099
- [26] —, "Tracking multiple moving objects for real-time robot navigation," 1100
Auton. Robots, vol. 8, no. 2, pp. 105–116, Apr. 2000. 1101
- [27] D. Schulz, W. Burgard, D. Fox, and A. Cremers, "Tracking multiple 1102
 moving targets with a mobile robot using particle filters and statistical data 1103

1104 association,” in *Proc. IEEE Int. Conf. Robotics and Automation*, Seoul, 1105 Korea, 2001, pp. 1665–1670.

1106 [28] B. Jensen and R. Siegwart, “Using EM to detect motion with mobile 1107 platforms,” in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Las 1108 Vegas, NV, 2003, pp. 1518–1523.

1109 [29] B. Jensen, R. Philippsen, and R. Siegwart, “Narrative situation assessment 1110 for human–robot interaction,” in *Proc. IEEE Int. Conf. Robotics and 1111 Automation*, Taipei, Taiwan, 2003, pp. 1503–1508.

1112 [30] B. Jensen, G. Froidevaux, X. Greppin, A. Lorotte, L. Mayor, M. Meisser, 1113 G. Ramel, and R. Siegwart, “Multi-robot–human interaction and visitor 1114 flow management,” in *Proc. IEEE Int. Conf. Robotics and Automation*, 1115 Taipei, Taiwan, 2003, pp. 2388–2393.

1116 [31] A. Hilti, B. Nourbakhsh, I. Jensen, and R. Siegwart, “Narrative-level 1117 visual interpretation of human motion for human–robot interaction,” in 1118 *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, Maui, HI, 2001, 1119 pp. 2074–2079.

1120 [32] B. Jensen, G. Froidevaux, X. Greppin, A. Lorotte, L. Mayor, M. Meisser, 1121 G. Ramel, and R. Siegwart, “The interactive autonomous mobile system 1122 robox,” in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 1123 Lausanne, Switzerland, 2002, pp. 1221–1227.

1124 [33] A. Drygajlo, P. Prodanov, G. Ramel, M. Meisser, and R. Siegwart, “On 1125 developing voice enabled interface for interactive tour-guide robots,” 1126 *J. Adv. Robot., Robot. Soc. Jpn.*, vol. 17, no. 7, pp. 599–616, Nov. 2003.

1127 [34] D. Spiliotopoulos, I. Androustopoulos, and C. Spyropoulos, “Human– 1128 robot interaction based on spoken natural language dialogue,” in *Eur. 1129 Workshop Service and Humanoid Robots (ServiceRob)*, Santorini, Greece, 1130 2001, pp. 1057–1060.

1131 [35] S. Young, J. Odell, D. Ollason, and P. Woodland, *The HTK Book, 1132 Version 3.0*. Redmond, WA: Microsoft Corp., 2000.

1133 [36] P. Renevey and A. Drygajlo, “Securized flexible vocabulary voice mes- 1134 saging system on Unix workstation with ISDN connection,” in *Eur. Conf. 1135 Speech Communication and Technology (Eurospeech)*, Rhodes, Greece, 1136 1997, pp. 1615–1619.

1137 [37] X. Huang, A. Acero, and H. Hon, *Spoken Language Processing*. Upper 1138 Saddle River, NJ: Prentice-Hall, 2001.

1139 [38] C. Breazeal and B. Scasselatti, “How to build robots that make friends 1140 and influence people,” in *Proc. IEEE Int. Conf. Intelligent Robots and 1141 Systems*, Kyonjiu, Korea, 1999, pp. 858–863.

1142 [39] B. Siebenhaar-Röllli, B. Zellner-Keller, and E. Keller, *Phonetic and Tim- 1143 ing Considerations in a Swiss High German TTS System*. New York: 1144 Wiley, 2001, pp. 165–175.

1145 [40] E. Keller and B. Zellner, *A Timing Model for Fast French*. York, U.K.: 1146 Univ. York, 1996, pp. 53–75.

1147 [41] T. Dutoit, V. Pagel, N. Pierret, F. Bataille, and O. V. der Vrecken, “The 1148 MBROLA project: Towards a set of high-quality speech synthesizers 1149 free of use for non-commercial purposes,” in *Proc. Int. Conf. Spoken 1150 Language Processing (ICSLP)*, Philadelphia, PA, 1996, vol. 3, pp. 1393– 1151 1396.

1152 [42] *IBM ViaVoice* [Online]. Available: <http://www-306.ibm.com/software/voice/viavoice/> 1153

1154 [43] L. Mayor, B. Jensen, A. Lorotte, and R. Siegwart, “Improving the ex- 1155 pressiveness of mobile robots,” in *Proc. Robot and Human Interactive 1156 Communication (ROMAN)*, Berlin, Germany, 2002, pp. 325–330.

1157 [44] P. Ekman and R. Davidson, *The Nature of Emotion: Fundamental Ques- 1158 tions*. New York: Oxford Univ. Press, 1994.

1159 [45] N. Tomatis, G. Térrien, R. Piguët, D. Burnier, S. Bouabdallah, and 1160 R. Siegwart, “Designing a secure and robust mobile interacting robot 1161 for the long term,” in *IEEE Int. Conf. Robotics and Automation*, Taipei, 1162 Taiwan, 2003, pp. 4246–4251.

1163 [46] P. Prodanov and A. Drygajlo, “Bayesian networks for spoken dialogue 1164 management in multimodal systems of tour-guide robots,” in *Proc. 8th 1165 Eur. Conf. Speech Communication and Technology*, Geneva, Switzerland, 1166 2003, pp. 1057–1060.



Nicola Tomatis received the M.Sc. degree in com- 1178 puter science from the Swiss Federal Institute of 1179 Technology (ETH), Zurich, Switzerland, in 1998, 1180 and the Ph.D. degree from the Swiss Federal Institute 1181 of Technology (EPFL), Lausanne, Switzerland, in 1182 2001. 1183 His research covered metric and topological (hy- 1184 brid) mobile robot navigation, computer vision, and 1185 sensor data fusion. Since autumn 2001, he holds 1186 a part-time position as Senior Researcher with the 1187 Autonomous Systems Lab. He is currently the CEO 1188 of BlueBotics SA, Laussane, Switzerland, which is a start-up involved in mobile 1189 robotics. 1190



Laetitia Mayor studied at EPFL and Carnegie Mel- 1191 lon University and received the master’s degree in 1192 microengineering from the Swiss Federal Institute 1193 of Technology (EPFL), Lausanne, Switzerland, in 1194 2002. In her master’s thesis, she developed a concept 1195 for emotional human–robot interaction. 1196 In spring 2002, she joined the *Expo.02* robotics 1197 team at EPFL to work on emotional human–robot 1198 interaction and the development of scenarios. After 1199 the successful completion of the *Expo.02* project, she 1200 joined Helbling Technik AG. 1201



Andrzej Drygajlo (M’84) received the M.Sc. and 1202 Ph.D. (*summa cum laude*) degrees in electronics 1203 engineering from the Silesian Technical University, 1204 Gliwice, Poland, in 1974 and 1983, respectively. 1205 In 1974, he joined the Institute of Electronics at 1206 the Silesian Technical University where he was an 1207 Assistant Professor from 1983 to 1990. Since 1990, 1208 he has been affiliated with the Signal Processing 1209 Laboratory (LTS) of the Swiss Federal Institute of 1210 Technology (EPFL), Lausanne, Switzerland, where 1211 he presently works as a Research Associate. In 1993, 1212

he created the Speech Processing Group of the LTS. His current research 1213 interests are man–machine communication, speech processing, and biometrics. 1214 Currently, he conducts research and teaching in these domains at the EPFL 1215 and the University of Lausanne. He participates in numerous national and 1216 international projects and is member of various scientific committees. He is 1217 currently an Advisor on numerous Ph.D. theses. He is the author/coauthor of 1218 more than 70 research publications, including several book chapters, together 1219 with his own book publications. He is also an appointed expert nominated by 1220 the European Commission in the domain of speech and language technology. 1221

Dr. Drygajlo is a member of the EURASIP, International Speech Commu- 1222 nication Association (ISCA), and European Circuit Society (ECS) professional 1223 groups. 1224



Roland Siegwart (M’90–SM’03) received the M.Sc. 1225 degree in ME and the doctoral degree from the Swiss 1226 Federal Institute of Technology (ETH), Zurich, 1227 Switzerland, in 1983 and 1989, respectively. 1228

After his Ph.D. studies, he spent one year as a post- 1229 doc at Stanford University, where he was involved in 1230 microrobots and tactile gripping. From 1991 to 1996, 1231 he worked part time as R&D Director at MECOS 1232 Traxler AG and as a Lecturer and Deputy Head at 1233 the Institute of Robotics, ETH. Since 1996, he has 1234 been a Full Professor for Autonomous Systems and 1235 Robotics at the Swiss Federal Institute of Technology, Lausanne (EPFL), and 1236 since 2002, a Vice Dean of the School of Engineering. He leads a research 1237 group of around 25 people working in the field of robotics and mechatronics. 1238 He has published over 100 papers in the field of mechatronics and robotics, is 1239 an active member of various scientific committees, and is a cofounder of several 1240 spin-off companies. 1241

Dr. Siegwart was the General Chair of IROS 2002 and is currently VP for 1242 Technical Activities of the IEEE Robotics and Automation Society. 1243



Björn Jensen (S’02–M’04) received the master’s 1167 degree in electrical engineering and business admin- 1168 istration from the Technical University of Darmstadt, 1169 Germany, in 1999. He is working toward the Ph.D. 1170 degree at the Autonomous Systems Lab (ASL), 1171 Swiss Federal Institute of Technology (EPFL), Lau- 1172 sanne, Switzerland.

His main interest is in enhancing man–machine 1173 communication using probabilistic algorithms for 1174 feature extraction, data association, tracking, and 1175 scene interpretation. 1176 1177

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