FROG

FUN ROBOTIC OUTDOOR GUIDE

Deliverable: D1.1

Functional Requirements, Interaction and Constraints

Consortium

UNIVERSITEIT VAN AMSTERDAM (UvA)

YDREAMS - INFORMATICA S.A. (YD)

IDMIND - ENGENHARIA DE SISTEMAS LDA (IDM)

UNIVERSIDAD PABLO DE OLAVIDE (UPO)

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE (ICL)

Grant agreement no. **288235**

Funding scheme **STREP**

SEVENTH FRAMEWORK PROGRAMME

DOCUMENT INFORMATION

Project	

Project acronym:	FROG
Project full title:	Fun Robotic Outdoor Guide
Grant agreement no.:	288235
Funding scheme:	STREP
Project start date:	1 October 2011
Project duration:	36 Months
Call topic:	ICT-2011.2.1 Cognitive Systems and Robotics (a), (d)
Project web-site:	www.frogrobot.eu

Document	

Deliverable number:	D1.1
Deliverable title:	Functional Requirements, Interaction and Constraints
Due date of deliverable:	M6 - 31 March 2012
Actual submission date:	M7 - 11 April 2012
Editors:	All Partners
Authors:	All Partners
Reviewers:	All Partners
Participating beneficiaries:	All Partners
Work Package no.:	1
Work Package title:	Robot Specification, Design and Construction
Work Package leader:	IDMind
Work Package participants:	All Partners
Estimated person-months for deliverable:	9
Dissemination level:	Public
Nature:	Report
Version:	1
Draft/Final:	Final
No of pages (including cover):	51
Keywords:	scenario, requirements, system specifications

Table of Contents

1. Summary

FROG proposes to develop a guide robot with a winning personality and behaviours that will engage tourists in a fun exploration of outdoor attractions. The work encompasses innovation in the areas of vision-based detection, robotics design and navigation, human-robot interaction, affective computing, intelligent agent architecture and dependable autonomous outdoor robot operation.

In the first months of the project the partners have been concentrating their efforts on the definition of the scenario, the envisioned interactions between FROG and the surrounding environment, interactions between FROG and humans and the required technology to accomplish it.

The partners organized and participated in an end-user week to get a better understanding of the global user requirements and the constraints involved.

This deliverable addresses the specification of FROG taking into account the performance needs in terms of mechanics, sensing, actuation, motion control, communications, processing and energy, to address the project's scientific and technological challenges.

2. Scenario Definition of FROG

FROG proposes the use of robotic characters in the exploration of outdoor historical and cultural sites. For the scenario definition of FROG, this section will start by describing the selected end-user's working environments. Then it will describe the end-user week that was organized in each of the selected sites. Finally it will present the envisioned use of FROG in those environments.

2.1. Working Environment

The application of outdoor guides is expected to impact on tourist sites. During the elaboration of the proposal, the Consortium contacted with some potential end-users who were very receptive to the possibility of having a fun interactive robot as a highly attractive and competitive advantage, and who were available to provide their facilities for the project tests. These contacts were re-established after the project kick-off, and after some discussions between the partners the consortium agreed on having two different environments for the pilot-experiments: the Lisbon Zoo and the Royal Alcazar in Seville.

2.1.1. Lisbon Zoo

The Lisbon City Zoo has been identified as a potential end-user and there were two meetings with them before the End-User Workshop Week.

The Lisbon City Zoo was founded in 1884 and its mission includes the development and promotion of a park (zoo and botanical) as a centre for conservation, breeding and reintroduction into their natural habitat of endangered species through scientific research and environmental enrichment programs. In this promotion, education is coupled with a strong element of fun and entertainment.

From the first contact with the Zoo, they have shown great interested in the FROG robot concept. They have a very strong content department with experience in creating workshops for children, for example, and are willing to collaborate with the Consortium envisioning ways of adapting or creating their content to the project purposes. They also want to be active collaborators in the dissemination of the project to the media and even seek for ways of having the robot as a permanent creature in the zoo after the end of the project.

Figure 1 overviews the Zoo park installation. An initial analysis of the scenario from the point of view of robot operation was performed. Some possible routes for FROG are highlighted in magenta. Figure 2 shows some pictures that have been taken along the highlighted route. A video from this scenario can be found in:

http://www.idmind.pt/documents/movies/videoZoo.mpg

Figure 1. Lisbon Zoo

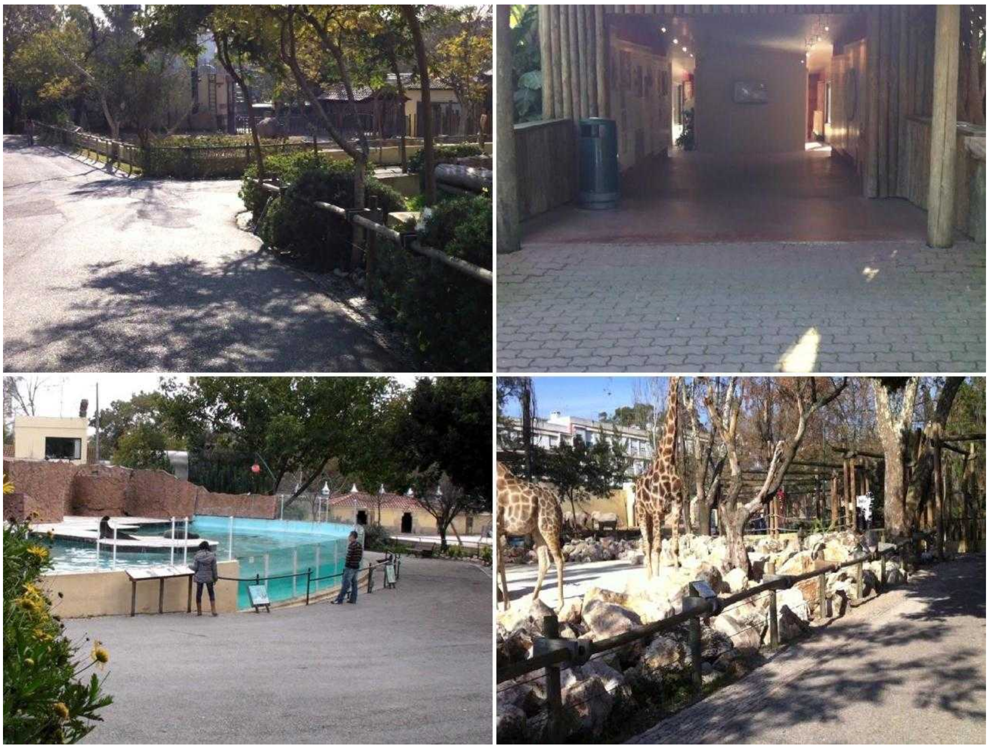

Figure 2. Pictures from the selected routes

2.1.2. Royal Alcazar - Seville

The Alcazar is a monumental complex located in Seville dating from the Muslim period. It consists of a series of palaces from different times and styles. It was originally a Moorish fort built in the 10th century. In fact, Alcazar is a Spanish word, synonym of fortified castle, that itself comes from the Arabic word *al qasr*, which means palace or fortress. The construction of the current Royal Alcazar began in the 14th century. This monument was declared a World Heritage Site in 1987. Figure 3 shows some pictures of the monument.

The Royal Alcazar has around 1.200.000 visitors per year. In meetings, the managers were very keen on the project concepts, and we have permission for data gathering and tests there.

An initial analysis of the scenario from the point of view of robot operation has been performed. Two videos of two different potential scenarios can be found:

http://www.upo.es/isa/lmercab/video/frog-scenario-1.m4v Outdoors scenario

http://www.upo.es/isa/lmercab/video/frog-scenario-2.m4v A mixture of outdoors and indoors.

Figure 3 shows the approximate shape of the routes shown in the videos. Figure 4 shows some pictures that have been taken from the highlighted route.

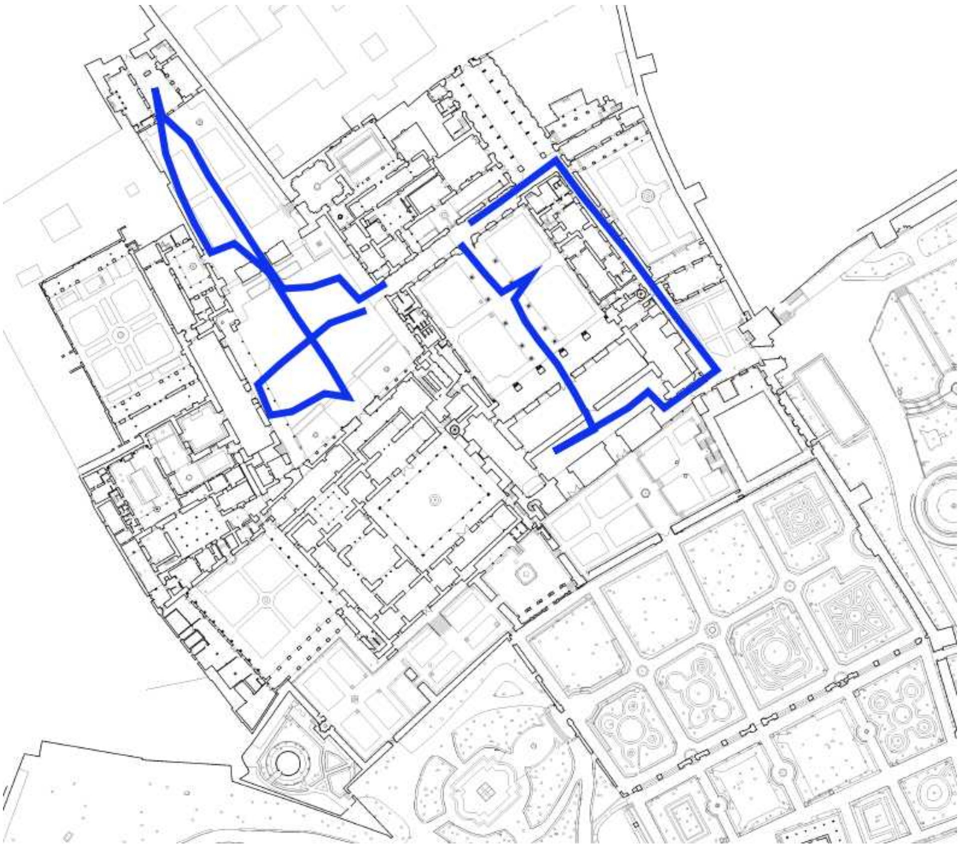

Figure 3. Approximate shape of the routes shown in the videos

Figure 4. Pictures from the selected routes

2.2. Visitor Experience

2.2.1. Summary of the End-User Week Workshops

In the week of March 4-10 an end-user workshop was held at both test sites of the FROG project. The workshop started in the Lisbon City Zoo where the consortium checked the places accessible for the robot, observed visitors and guides and interviewed two guides.

The second part of the week was in the Royal Alcazar of Seville where the consortium also checked the accessibility for the robot and observed visitors and guides and interviewed several guides. During the workshop in the Royal Alcazar the consortium mapped the visitor experience and with that information set the basic scenario. Also the consortium had a meeting with the manager of activities of the Royal Alcazar and the management board of Italica Roman City near Seville to confirm the ideas and verify that no important visitor experiences were missing, and to confirm what their first thoughts about the opportunities of the robot.

In the following paragraphs the results of the end-user workshop will be given concerning the visitor experiences, the user requirements and the basic and advanced scenarios.

2.2.2 Research Visitor Experience

The goal of this research is to give insight into the visitor experiences in tourist sites and the factors that influence the visitor experience. From this study it will become clear which parts of these visitor experiences can be improved; and thus these results will give insight into where there is room for human-robot interaction with a robotic guide and what the robot functions can be.

2.2.3 Methods

The visitor experience data is divers, and obtained from observations, interviews and a workshop. To collect data, 2 researchers followed 4 tours at the two test sites and interviewed the tour guides afterwards. The researchers also observed behaviour of visitors without guides at the two sites. Finally, to have broader support for the findings on visitor behaviour and experience, a workshop was organized with project partners (who had the opportunity both to go around by themselves and to be guided through the sites) on their own visitor experiences.

The observations of the guided tours were videotaped by one of the researchers, and notes were taken by the other researcher. When the tour was finished the guide was interviewed and answered some questions about visitor experiences:

What is the purpose of your tour?

What is the aim of the tour?

What is the main exhibit?

Do you notice differences between groups? How do you deal with that?

Is guiding children different from guiding adults? If so, in what way?

How do you gain and keep the attention of the visitors?

What would you like to change to improve the visitor behaviour?

Any other comments?

After the interview both researchers completed their notes about the observation and the interview. The observations of visitors going around on their own were noted by both researchers (both were in different spots for most of the time). All these notes and then results of the workshop were input for the visitor experience map.

The workshop on visitor experiences organized for the project partners, was performed by two researchers. One of them lead the workshop, the other one took notes and summarized the findings. These findings were used in a discussion, directly following the workshop, that was meant to further define the scenarios for the FROG robot. The workshop was videotaped to have the opportunity to review the remarks of the project partners afterwards.

The workshop consisted of different parts. First the project partners were asked to write down their experiences, feelings, remarks and observations about the two sites while being guided or not being guided (for these two conditions two different colours post-its were used, pink for being guided and yellow for walking around by themselves). Then all post-its were collected and during a group discussion the post-its were ordered and clustered.

The results from the workshop were analysed and the clusters were named and reordered if necessary to obtain clusters that covered the experiences of the project partners (for this the video was watched carefully to find meaningful remarks that were lost during the workshop). This resulted in the visitor experience map (see Figure 5) which gives visual information about the factors that influenced the visitor experiences. Six main clusters and 15 sub-clusters and lots of connections can be found in the graph. This map will be explained and commented further in the next section.

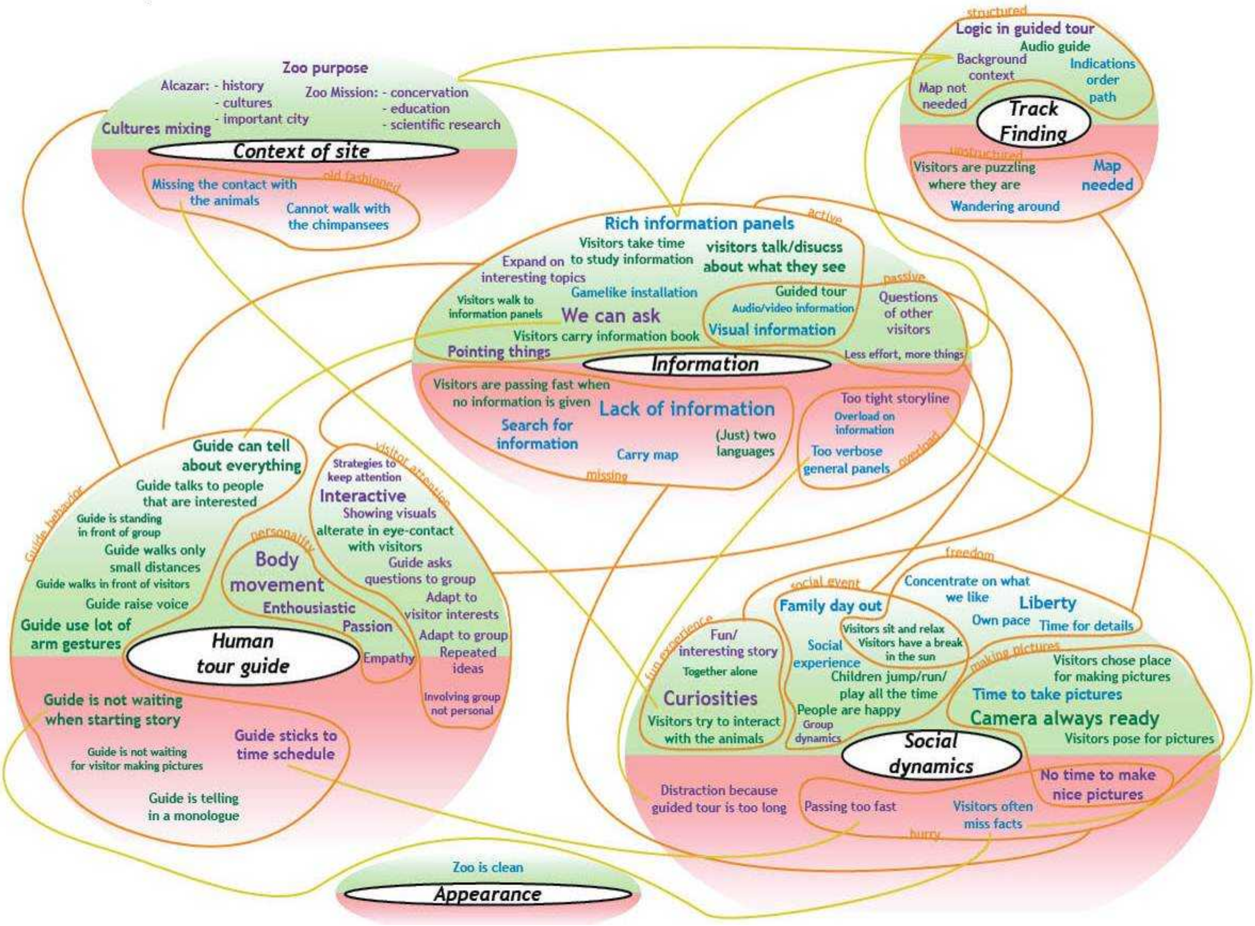

Figure 5. Visitor experience map

2.2.4 Results Visitor Experience

When visitors are in tourist sites such as the Zoo (in Lisbon) or the Royal Alcazar in Seville, they are always looking for information, because this is one of the fun experiences. That is why the cluster about information is placed in the middle of the scheme. All other clusters (except for appearance) are related to the information cluster. In the following paragraphs the factors that influence the visitor experience and their relation to each other will be described.

Explanation of colour codes used in the visitor experiences map:

- Black and white written words are names of the main clusters.
- Blue terms are visitors experiences from non-guided visitors.
- Purple terms are visitor experiences from visitors that followed a guided tour.
- Green terms are visitor experiences abstracted from observation.
- Green area of a cluster is positive, red is negative.
- Sizes of the words show the importance (bigger is more important).
- Orange circles cluster similar experiences in a secondary cluster.
- Oranges lines relate secondary clusters to each other.
- Yellow lines cluster related experiences to each other.

Context of site

The cluster context of site is the first cluster to be described, because visitors do know on forehand what kind of tourist site they are visiting and in what kind of context they will end up. But as we learned, the background information they have about the context is not always up to date and certainly not complete. So visitors are interested to find more information about the sites. Getting this information and having an entertaining day together are two of the main visitor experiences.

To have an idea of the information visitors like to have about the backgrounds of the sites, the background contexts of the Zoo and the Royal Alcazar will be described. The Zoo has three main missions, namely education of children and adults, conservation of species, and doing and giving others the opportunity to do scientific research. Other than most adults can remember the animals no longer live in small cages, but in enriched environments that resemble their natural environments. There the animals can show and practice their wild behaviour. Interaction with the animals is discouraged by the zoo. In the Royal Alcazar it is important to know about the history of the place, about the Christian and Muslim cultures that were going together for ages and about the importance of the harbour of the city of Seville in previous centuries.

Social dynamics

When visitors are going around a site (with or without guide) they do experience the site in a particular way. The zoo especially was experienced as a (family) day out, visitors went there and took a walk with family, friends and children and in the meantime they talked about everything, including the animals every now and then. In the Royal Alcazar visitors were searching for information more obviously, but having time to relax in the beautiful gardens also made this visits a social experience.

The group dynamics that were observed were interesting. In the Royal Alcazar couples of all ages, couples with children (mainly older than 10) and school classes were found, who looking at exhibits, discussing them and taking time to relax in the gardens. In the zoo often families consisting of parents, grandparents or siblings of the parents and one or two children were observed, as well as young couples without children and tourists. These visitors were walking around and enjoying the weather and environment. When a group with children was at the Zoo, the whole visit was about the children having a fun time. The children were jumping, running and playing all the time and parents often pointed out animals for them. Also the children were asked to pose for a picture.

The smaller groups enjoyed having freedom in the sites to visit at their own pace and have time to concentrate on what they like. In that way they have time for details if they want, and time to take a break to have more social interaction. In small groups visitors do not always have the same interests, but they give each of the group the freedom to look for what they want, and they go on when all members of the group have had the opportunity to consume the amount of information they want. This liberty to go anywhere is not experienced when being guided, this is one of the reasons why people choose not to have a guided tour. Having this freedom is related to track finding, because the visit is not always structured without a guide.

The number of cameras you will find in tourist places is remarkable. All groups of visitors carried at least one camera, most of the time visitors had small, easy to use, digital cameras, some had a more elaborate camera. Most of the visitors not following a guided tour had their cameras ready and took pictures of every exhibit they see. Visitors also often posed or let friends or their children pose for the camera. The ones with more elaborate cameras were searching for the best place to take nice pictures of the exhibits. When visitors followed a guided tour there was less time to take pictures, or the visitor would fall behind.

The sub-cluster fun experience is closely related to the information cluster. Listening actively to the guide telling funny/interesting stories about the site and curiosities about the animals was one of the fun experiences for visitors, because these stories cannot be obtained anywhere else. And these curiosities are part of information visitors will remember best and tell to their friends at home (as proof that they have visited the place). This curiosities can be for example the names of the giraffes or the fact that the Portuguese were the first to bring the rhinoceros to Europe.

Another part of the fun experience in the zoo is interacting with the animals, which is discouraged by the zoo, because it stresses the animals. By having the animals in rich environments visitors have the chance to see the animals as they are in the wild. No interaction at all with the animals is not achievable yet, but the situation is already much better for the animals than before (when people could walk with chimpanzees). In this the visitor finds part of the mission of the zoo, for which it is nice to have the background information to understand the context.

Visitors had negative social experiences while being guided. On the one hand a guided tour is long and the visitors got distracted in the end because they were overloaded with information. But on the other hand the guide always went on too fast so that visitors could not have a proper look at the species, fell behind if they wanted to take pictures of the animals and often missed things because the guide had already started talking while the group was not yet complete.

Information

The information visitors can consume depends on their own willingness. The information given at the sites is very different depending on whether visitors follow a guided tour or just walk around on their own. The amounts of written information in the Zoo differs from the information given in the Royal Alcazar. In the Zoo there is an information panel for all species. On this panel four main topics are discussed. In some areas there is more information about the species, for example at the tiger house and the temple of primates. Besides that, the visitors can go to a dolphin show, see how some of the animals are fed or have a guided tour.

In the Royal Alcazar only one short piece of text is placed in each different room. No additional information is given at other exhibitions. In the Royal Alcazar visitors can choose for an audio tour or a guided tour. In both sites visitors used maps to find out where they were and where they needed to go. Especially in the Royal Alcazar visitors were also carrying information books.

Figure 6. information panel Lisbon Zoo

Figure 7. Information panel Royal Alcazar

Visitors can obtain their information actively and passively, both are possible during wandering around and a having guided tour. Unguided visitors who were active in consuming information searched for the information panels, they read these panels, carry an information book brought from home and read it aloud, pointed at things, talked and discussed what they saw, took time to study the information and the exhibits. These active visitors liked the rich information panels in the zoo, and consumed the audio and visual information. When being guided visitors who actively obtained their information, asked the guide questions and listen carefully to his or her explanations.

Obtaining the information more passively was also possible when being guided and not being guided. These visitors were at the site for the most social event and did not search for information. When the information was given they looked at it briefly. Visual and audio information are easy to consume. When this kind of visitor was following a guided tour, he was mostly at the back of the group and listened to the guide and the questions other visitors asked. One of the fun experiences is getting information and learning about curiosities, so even when visitors were passive information consumers, they liked the fact that they obtained this information. For this kind of visitor following a guided tour means less effort, while more information is obtained.

When walking around on their own, visitors may have had the feeling that they could not consume too little information. In the Zoo and in the Alcazar as only one small information panel in just two languages per species or room was given. Visitors often needed to search for these information panels and their need for information was not satisfied. In the Royal Alcazar there was no more written information given, visitors could choose for an audio tour or read in an information book brought from home. In the zoo some more information about the tigers and the primates was given in the tiger-house and the temple of primates, which were specially designed for showing more information. Positive on these information houses is that visitors saw these information panels in the Zoo as very rich and the visual presentation of the information was very well experienced. But at the same time visitors could become overloaded by the rich information panels and the other information given in the compact spaces.

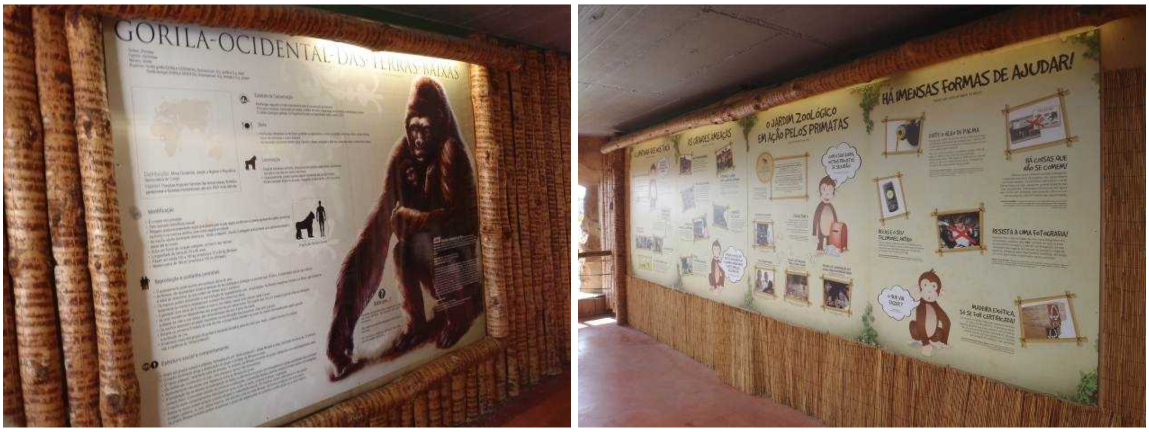

Figure 8. Examples of rich information panels in the temple of primates in the Lisbon Zoo.

At both sites visitors on guided tours received a lot of information. The guide often expanded on visitor interests, told about curiosities and pointed at remarkable things and visitors could ask questions of the guide, but the amount of information given is not always satisfying the visitors. Because a tour only lasts for a maximum of two hours, the guide has not enough time to tell everything he or she knows. The information is given in a short amount of time and the storyline is tight. Visitors had the idea of being overloaded by information, while they did not have enough time to have a proper look at the animals or small exhibits (such as remarkable tiles) other visitors were standing in front of at the same time.

Obtaining too much and too little information were both experienced negatively, but obtaining too much information was liked better, because visitors could then choose for themselves whether they had enough. When the information was not available at all the visitor was disappointed.

The above mentioned clusters (social dynamics and information) seem to have a contradiction in them. Visitors liked to go around and take pictures at their own pace, but then they experienced a lack of information both at the Zoo and at the Royal Alcazar. When guided around during a tour, visitors liked the information they received, especially the curiosities, but at the same time they did not like the speed of the tour and the tight time schedule and tight storyline. Finally a guided tour was experienced as better than wandering around, because visitors that fell behind could always follow some parts of the story later in the tour, and the tour only lasted for one and a half to two hours. Afterwards visitors could go and visit the site at their own pace and get more social again.

Track finding

Visitors liked the structure of guided tours or a path that indicated the route and leads them through a site in a logical order, having background context and curiosities at the right time and at the right place. Knowledge about the background context can help to structure the visit, even if the background context is obtained by following an audio tour or using books. For visitors it was positive that they did not need a map and did not have to puzzle where they were. Because both the Zoo and the Royal Alcazar are not structured in a way that a clear path is given there is a risk that visitors wander around, missing a lot of the information.

Human tour guide

The guides that gave the tours have certain personalities. All four guides in this research were enthusiastic, told stories with passion and used a lot of body movement. These personalities are important to gain the empathy visitors feel for the guide.

The guides had some behaviour and used different strategies to keep the attention of the visitors. The main strategy was to interact with the visitors. By keeping the guided tour interactive, visitors were engaged more than when the guide delivered a monologue. More strategies were: making and keeping eye-contact with the visitors, showing visuals, asking questions of the visitors, giving room for visitors to ask questions and repeating ideas. Where the last factor is partly negative, because repeating the idea helps to remind the background information, but visitors also do have the idea they hear it over and over again. Another negative factor is involving one person; when the guide is speaking too long to one person, he feels embarrassed, but the guide alternating the attention is experienced as positive.

The guide could also adapt to the group, because no two groups behave the same. Especially when guiding children, the tour needed to be different. The guide could adapt to the group interest by responding to questions and adapting the content of the tour. But depending on the group the route through the site could also be changed. This is no problem for the guides participating in this research, because they have enough experience to do this. For children, changing the subject to something the children would pay attention to was a useful strategy. Sometimes children get distracted by something around them. The guide knows all the stories and could shift his subject, to keep the attention of the children.

The behaviour of the guides showed some common aspects. As mentioned before, the guides could all tell freely about everything they encounter and had no problems in answering anything the visitors ask. A guide walks a bit in front of the group. This made the visitors move and gave the guide the time to prepare at the new exhibit. And the guide made sure the distances between two exhibits he wanted to talk about were small. When arriving at a new exhibit, the guide did not wait for the group to be complete, but started to talk to the visitors that were already close and that were interested in the story. If the guide wanted to tell something all visitors should hear, he raised his voice. The guide did not use his authority for adults, because these visitors can choose for themselves whether they want to pay attention. When guiding children the guide usually teamed up with the teacher and the teacher would use his or her authority if necessary.

When the guide talked about an exhibit, he was in front of the visitors and the visitors grouped up in a sort of semi-circle. The guide used a lot of arm gestures while telling the story, and the guide often pointed at the exhibit. Mutual gaze to the exhibit is important for the visitors to start looking at the exhibit.

A negatively experienced property of guided tours was the rush. The guide needs to stick to a tight time schedule. When a guide lost time at one part of the site (e.g. because of visitors asking questions), the guide needed to win time at another part of the tour. The strategy used was that they only mentioned the name and one or two curiosities and then walked on. This rush meant that visitors did not have enough time to take a proper look at some of the exhibitions and the visitors did not have time to take photographs. As the guide started talking before the group is not complete, visitors often missed facts.

2.2.5 Infographic Visitor Experience

In Figure 9 a visual overview of the results of the visitor experiences is presented. The differences between being guided and wandering around are made visual. The experiences for the viewer of the infographic will be the same as for the visitors in the tourist sites. On the side about being guided, the visitor follows a path, the information is structured and of similar amount for each exhibit. The experience is clear and full of information.

On the other side; wandering around, the experience is more chaotic, but also more relaxed and more fun, definitely making time for pictures. But less information is given, except for one place where a large amount of information is given.

Figure 9. Infographic visitor experience

2.2.6 Discussion

The visitor experiences at the Zoo and in the Royal Alcazar are complex to grasp entirely and there are clear differences between the sites. The visitors that came to the Royal Alcazar and the Zoo differed. The aim of visitors in the Zoo was more to have a social event than to search for information. For visitors of the Royal Alcazar, getting information was more important. In the end a robotic guide will probably have more added value in the Royal Alcazar, because at the Zoo visitors will find their way and they will be satisfied to have only the animals to look at. This will be especially true for the children.

Despite the differences mentioned above, the main visitor experiences showed some similarities as we can see in the map where they are put together. Most negative factors found at both places about wandering around on your own were about the lack of information, while the negative factors for being guided were about the speed the guide had to finish the tour.

Differences in the visitor experiences will influence the user requirements of the site, but by using the experiences that are the same, a consistent set of user requirements can be made.

2.2.7 User Requirements

The user requirements are obtained from the positive and negative factors in the graph. Only factors that influence the visitor experience in both sites have been used to determine the user requirements.

Context:

- The mission should be given as basic information/background context

Visit experience:

- Visitors should have time to take pictures
- Visitors should have time to go at their own pace
- Children should have time to play
- The information should contain curiosities/fun stories about sites
- The information given should be based on the visitors' interests
- The information given should be adapted to visitors' routes

Information:

- Obtaining basic information should not require much effort from visitors
- Visitors should be able to choose the amount of information they want to consume
- The information should not overload the visitors
- The information must be given in different languages (at least English and the language of the country)
- Visitors should have time to study the information
- Visitors should have space to study the information

Track finding:

- The guide should show people where they are
- The guided tour should structure the visit

Guide:

- The information given by a guide should not be a monologue
- The guide should engage all people in the group (not focus on one)
- The guide should adapt to visitors' interests
- The guide should show visuals/augmented reality
- The guide should lead a group through the site in a structured way
- The guide should be able to point at things
- The guide should be able to give information about everything (if visitors ask for it).

2.3. Use Case Definition

Based on both the aims of the proposal and the experiences acquired during the End-User Workshops, FROG's use case scenario has been defined by adopting an agile approach. The project will start by developing a basic interaction scenario that will be enriched iteratively and incrementally with more complex features. Although the pilot tests will be carried out in two completely different environments, the difference between the two is just in content rather than in the interaction scenario, and the robot can be easily tested in both environments.

2.3.1. Basic Interaction Scenario

The flowchart of the Basic Interaction Scenario is depicted in Figure 10. Each of the states of this flowchart can be described as follows.

Time out - reach next way-point

The robotic guide starts to drive around from one waypoint to the next waypoint following a default circuit. During this process the robot is always detecting obstacles and avoiding collisions. While driving from one waypoint to the next the robot is looking for visitors. When no visitors are perceived, the robot continues in *Time out – reach next way-point*. If the robot detects (a group of) visitors, the robot continues at *Detect visitors*.

Detect visitors

When the robot recognizes (a group of) visitors (maximum distance is 25 meters), the robot starts to approach the visitors. From a distance the robot is detecting their pose, so that it can approach the visitors from the front if possible. Based on the visitors' pose, the robot reasons on whether the visitors are engaged in a conversation or are facing the robot. Once they are facing the robot, it *continues* by introducing itself at *Introduce robot.*

Introduce robot

The robot stops close to the visitors and introduces itself (by using screen/sound/speech/recorded speech/movie/augmented reality; still an open issue). If the visitors seemed (at *Detect visitors*) to be engaged with each other, the robot offers to guide them (continue at *Guide visitors*). If the visitors were (at *Detect visitors*) engaged by an exhibit close by, the robot offers to tell something about the exhibit. To give an answer to the suggestion of the robot, visitors can nod yes or shake no with their heads. (It might be problematic if different visitors in one group answer differently). If the answer is yes, the robot continues at *Tell about exhibit*; if no, the robot continues at *Guide visitors*. If the robot does not receive a reaction within a minute, the robot continues at *Time out – reach next way-point*.

Tell about exhibit

The robot detects whether the visitors are still close. If not, the robot waits and checks if the visitors are getting closer. If the visitors are not getting closer the robot will continue at *Time out – reach next way-point*. If yes, the robot starts a story (and tells at least one curiosity) about the exhibit. To tell the story (the way it tells the story is still an open issue) the robot can point at the exhibit, use pictures, movies and/or augmented reality to clarify the story. During the story the robot detects the facial expressions. If visitors seem to be bored, the robot sticks to the basic information and after that the robot continues at *Guide visitors*. If visitors seem to be engaged, the robot asks the visitors if they want to hear more. The visitors can react with nodding or shaking their heads. If not the robot continues at *Guide visitors*, if yes the robot will tell a more elaborated story. After the story has finished, the robot continues at *Guide visitors* (with a maximum of three exhibits, after that the robot will leave, continuing at *Leave visitors*).

Guide visitors

The robot asks the visitors if they want to be guided. The visitors react with nodding or shaking the head. If they do not want to be guided the robot continues at *Leave visitors*. If yes, the robot asks the visitors if they have already seen exhibit “x” (next way-point). To answer the question of the robot visitors can nod or shake with their heads. If yes the robot suggests the next exhibit (following next way-point). If yes again, the robot continues at *Leave visitors*. If no, the robot asks the visitors to follow and then the robot will guide the visitors to the next exhibit. At the exhibit the robot stops and continues at *Tell about exhibit*. If the robot does not receive a reaction within a minute, the robot continues at *Time out – reach next way-point*.

Leave visitors

If visitors do not want to be guided, have already seen some exhibits (as they indicated to the robot in *Guide visitors*) or have been guided through three exhibits, the robot will leave the visitors. The robot thanks the visitors for their attention and wishes them a pleasant stay at the site. Then the robot turns around and leaves the visitors, detects where it is and continues at *Time out – reaching next way-point*.

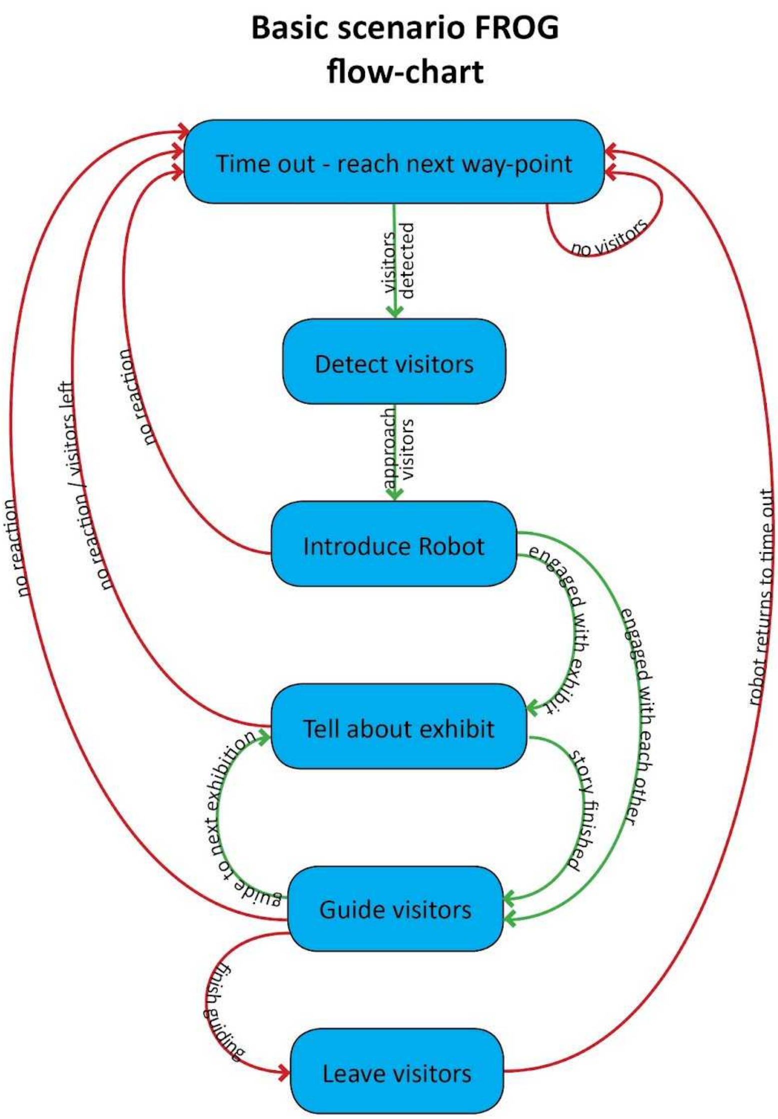

Figure 10. Flow chart of basic scenario

2.3.2. Advanced Scenarios

The consortium will focus on the basic scenario which covers the different expertises from the Consortium partners and fulfils the proposed goals of the project. During the project development there might be some deviations from this envisioned scenario. Some features might require some corrections and some others might be added. The advanced features that were discussed by the Consortium, some of them clearly beyond the scope of the project, and that will be included in the FROG scenario only if time permits, can be summarised as follows.

Recognizing visitors that were guided before

In this advanced scenario the robot is able to recognize (groups of) visitors that it has already been guiding or that have refused guidance before. If a group actively approaches the robot then it should be possible to get guidance (again). With this knowledge the robot will be able to decide not to approach groups that it has already lead through exhibits. It would be even more interesting if the robot could decide not to approach visitors in the range of exhibits that they already have been guided through, only approaching them in the range of exhibits that they have not been guided through (as the maximum of guided exhibits is set to three at a time). In this case, the robot should not give the whole introduction again, but it should just greet the visitors and offer to guide them again.

Recognizing if visitors do not react because they are taking pictures/movies

In this advanced scenario the robot will not go away if visitors do not react because they are taking pictures. When visitors are taking pictures of exhibits the robot is not able to detect their faces. The robot may also recognize that the visitors are taking pictures by their pose and in that case wait a bit longer or ask the visitors to inform him when to proceed (e.g., the robot asks the visitors after a while if they are ready with the pictures, then the visitors can react with nodding or shaking the head or pushing a button on a touch screen saying “proceed with the guided tour”).

Engaging children

Children are different in their behaviour from adults. They are often moving, jumping, running and playing. When children are engaged in the right way, they can pay attention to the content the robot is giving. It is most likely that children will be curious about the robot and that they will try to interact with the robot in a sense that has nothing to do with the guiding functionality of the robot. The robot should be aware of the presence of children and have special interactions for them, besides the guiding functionality for the group. This can be in the form of games on the screen or assignments in real life. For the possible interaction between robot and children, the guiding behaviour of a human guide when guiding children must be studied.

Ask more explicit feedback about the content the robot will be showing to the visitors

In the basic scenario the robot differs in the amount of content provided, based on the recognized facial expressions (bored or engaged). This might not always be a good reflection of the situation, therefore the robot should be able to ask visitors more explicitly about what they want to know. The visitors should have sufficient ways to make this interaction as natural and pleasant as possible.

Follow a group of visitors to their next way-point and tell about that

In the basic scenario the robot will lead the group from an exhibit (way-point) to another exhibit. The robot goes in front and guides the visitors through part of the exhibition. The other way around might also be an option: in that case visitors ask the robot to follow them to a specific exhibit from which they want to get information and in this case the robot follows the group. Arrived at the exhibit, the robot recognizes the place and will start to tell a story about the exhibit. After that, visitors can ask the robot to follow again, or the robot will be left behind.

Recognize if animals are visible

One of the test-sites for the robot is the Zoo. If the robot has a default circuit to drive, the robot will go along animal houses where possibly not all animals are visible. We learned from the human tour guides that they (especially when guiding children) do not mention the animals that are not visible. They just pass by the cage and go to the next exhibit. It would be nice if the robot (similar to the human tour guide) is aware of the presence of the animals. So visitors will not be bothered with a story about animals that are not there, and engagement will be higher if the robot only tells the interesting parts.

3. Identification of Ethical Issues

The FROG project is susceptible to ethical issues, because the robotic guide will have interactions with humans, detect the users and will collect vision data about them. Also, the user studies that will involve human subjects who will be exposed to early versions of the robot may need to consider ethical issues of privacy and deception. For these interactions only healthy humans who fully consent for their contribution to the research will be used to participate.

The ethical issues can be about the robot and its hardware and software uses, or the user studies that will be performed during the research. For the remainder of this section, the researchers have identified possible ethical issues that may arise. The ethical board of the FROG project will be asked to evaluate our usage scenarios and to identify any further ethical concerns we may have missed.

Ethical Issues:

1. The robot will save some information when recognizing and re-recognizing visitors. That the robot collects this information about visitors might be a problem, because visitors may not know this and may not consent to this.

   The recorded data will only be used for the robot and its actions. The data collected will not be used for other purposes than the FROG research and will not be given to third parties (except in extreme cases as for police research). The user studies will include detailed consent forms with information on privacy issues and personal data collection, each experiment will have a debriefing component to make sure participants are aware of what they have consented to.

2. The robot can make an error while detecting poses or facial expressions, visitors might draw conclusions about the robot's behavioural possibilities that are not met (e.g. knowing how the robot will react to visitors that are trying to hug it, or want to play soccer with it, or try to ask difficult questions). The robot could frustrate the users or make them feel uncomfortable or disappointed when it does not respond to them when they try to interact with it, or interact with them when they prefer not to, or not recognize when it approaches them a second time etc.

   If the robot recognizes that people are no longer amused with the presence of the robot (e.g. persons are frustrated, upset or frightened) the robot will leave the visitor. If the robot is not sure about the emotion the visitor shows, but there is no positive response, the robot will leave. In study setting, the human subjects will be debriefed to explain any concerns the subjects may have had.

3. The robot can fail in navigation and object detection and collide with visitors or drive into the wrong place (e.g. the bushes).

   Different levels of safety will be ensured in the robot's hardware and software. The robot will stop automatically when it is outside its boundaries. Furthermore the maximum speed of the robot should be low, and in study settings a remote control emergency stop by an observer can be applied so a collision with visitors or exhibits will not occur.

4. During studies on location. Visitors must be aware that they are participants in a study. Therefore, before interaction with the robot, consent must be asked. The robot should at least clearly display information about the study so that visitors are aware they are participating.

   The visitors will be asked for consent to use their data for the research. The visitors can decide to stop participating at any time without any consequences for the visitor (until the *data analysis has started). The consent form will be handed to the participant before the interaction with the robot. The consent forms will address:*

- *Fully understanding the purpose of the research*
- *Participation is voluntary and can be stopped at any time*
- *The researchers videotape/audiotape/take notes*
- *The collected data will be anonymous and confidential*
- *Pictures/video or audio data may be used for publication*
- *Agreement to take part in the research referred to.*

5. The data collected for the research will be treated with the utmost care and the privacy of the participants will be protected.

   Data will be anonymised immediately and only aggregated results will be used for analysis. Collected data sets can only be made when the participant fully understands in what way the data sets will be made available and consent to this.

6. The UvA/UT has extensive experience in doing user studies (in the field of human-robot interaction)

   Internal ethical review procedures and protocols will be adhered to the design of each study.

4. System Specifications

Based on the specified scenario (environment and envisioned use cases) this section points out the overall system specifications. It starts with an overview of the current state-of-the-art and then it describes the main specifications in each of the involved fields of expertise.

4.1. State-Of-the-Art Review

4.1.1. Robust Operation of Robots in Outdoor Environments

Robustness is still a major issue in robotics, especially for robots employed in everyday human environments and in interactions with humans. Increasing complexity of today's robotic systems has raised the need for introducing robustness measures, which have to be fully integrated in the development process of a robotic platform. The robustness of a system is intrinsically related with its dependability, which is a concept already in use in the IT industry with the following attributes (Avizienis et al., 2001): availability in respect to readiness for usage; reliability in respect to continuity of service; maintainability to undergo modifications and repairs; safety in respect to avoidance of catastrophic consequences on the environment; security in prevention of unauthorized access and/or handling.

When building an outdoor robot platform, under the constraint that it will have to work in an unstructured / human environment and interact with humans, there is a set of challenges to think about: adequate robot chassis and locomotion to overcome the terrain where the robot evolves (Borenstein et al., 1996, Siegwart et al., 2004); resistance to unfavourable environmental conditions (e.g., dust and water); adequate onboard power and inclusion of an external charging station somewhere in the working environment; ensuring safety, for the humans, for the environment and for the robot itself.

In engineering, fault-tolerant design, also known as fail-safe design, is a design that enables a system to continue operation, possibly at a reduced level (also known as graceful degradation), rather than failing completely, when some part of the system fails. In the last twenty years, scientists and engineers have studied the problem of self-diagnostic and fault detection on mobile robots.

Nikam et al. (1999) presents an intensive self-diagnostic scheme, allowing a quick intervention from humans in solving the detected problem. Many articles have referred to self-diagnostics and fault detection to reduce the time of repairing or to operator fault alert. Lately, the critical requirements associated with robotic exploration of space and also deep sea, showed the importance in having self-repair or self-reorganization automatisms, allowing the robot to continue with its task. Hafbaur et al. (2007) presents a way to retain the functionality of a mobile robot after the presence of faults in the hardware through system reconfiguration. Other examples (Bongard et al., 2006), show legged robots that after losing the possibility to use one of the legs, recalculate the kinematics to be able to continue their mission without the damage/removed leg. Duan Zhuo-hua et al. (2005) presents an interesting state-of-the-art of Fault Diagnostics (FDD) and Fault Tolerant

Control (FTC) for wheeled mobile robots under Unknown Environments, providing an overview of the strategies that can be used to deal with robot hardware faults, including multiple model based approach, particle filter based approach, sensor fusion based approach, layered fault tolerant architectures and others.

The FROG project approaches this problem with the inclusion of a group of low-level (and low-cost) sensors associated with the robot basic functions (energy, locomotion, self-preservation, safety, etc) that are ruled by dedicated low-level control loops. Low-level information is passed to high-level behaviour trees, but at the same time the low-level control reacts to changes that can affect the robot operation, which is fundamental to the improvement of the overall system dependability.

Beyond the state of the art

The implementation of low-level safety measures running independently of high-level algorithms, complemented by onboard low-level self-diagnosis and fault detection, integrated into the robot high-level behaviour trees will go beyond the state of the art in terms of system dependability improvement. Providing the required robustness for a robot that will work autonomously in real outdoor environments populated by humans with whom it will be interacting.

Performance/criteria indicators

The robustness of operation will be evaluated in the following two ways. Low-level control loops, associated with robot basic functions (energy, locomotion, self-preservation and safety) will be evaluated based on their capacity to maintain their normal operation in presence of disturbances, and to generate adequate fault diagnosis in case of error. The overall system robustness will be measured by the capacity of high-level behaviour trees to generate adequate response in the presence of low-level fault diagnosis.

4.1.2. Robot Navigation and Localization

Localization and navigation in outdoor tourist scenarios will require maps with 3D geometric information and other information [Mirats-Tur et al., 2009]. Highly accurate, 6-DOF localization will be required for navigation and augmented reality applications. Furthermore, the FROG robot will have to be deployed in different new scenarios, and therefore, a map building phase, including relevant information for augmented reality, will be required. In FROG, off-the-shelf sensors will be mainly used. Vision, both monocular and stereo, will be the sensors employed for localization and tracking. Currently, systems able to perform vision-based SLAM in large scenarios of hundreds of meters and several thousands of features are available, both with stereo-vision [Paz et al., 2008] and monocular cameras [Caballero et al., 2009], [Strasdat et al., 2010] [Klein and Murray, 2009].

Nowadays, appearance models, like bag-of-words models, are employed in recognition systems and will be used for localization. Appearance-based localization and SLAM systems have been demonstrated at very large-scales [Cummins and Newman, 2009]. Metrics and appearance have been combined and demonstrated in outdoor scenarios, as in [Botterill et al., 2010], in trajectories of the order of several kilometres.

In FROG, the scale of the scenarios is lower than that is dealt with for current systems. The main concern is how the state-of-the-art methods behave in crowded scenarios, when lots of moving elements (persons) are present. Then, in FROG, 6DOF SLAM (for the deployment phase) and localization algorithms based on vision will be analysed under the circumstances of crowded outdoor scenarios. Bag-of-words-based location recognition and metric schemes based on vision and laser range-finders will be combined for mapping and precise localization. Moreover, FROG will use these results for the augmented reality tools. Finally, FROG will explore the combination of these with new sensors such as Time-of-Flight (ToF) cameras.

Several successful robots have been deployed in crowded scenarios during the last decade, like the ones described in [Burgard et al., 1999] for museum guiding, or [Siegwart et al., 2003]. These robots usually navigate in crowded places by computing the shortest path to the next goal and using local reactive navigation to reach the goal. However, none of these systems try to make the robot move in a human-like manner or follow the natural flow of people through the environment. Moreover, navigating in crowded outdoor scenarios poses additional challenges [Sanfeliu et al., 2010]. Navigating in a human scenario involves not only safety, but also social interaction and social awareness. For instance, robots must avoid socially unacceptable paths. The same consideration is applicable to all levels of the navigation stack, from decision and planning to low level navigation.

Several authors have considered human-awareness during task planning. In [Cirillo et al., 2009], a task planner considering humans is described. The robot planner builds on a plan recognition system that recognizes potential future plans of the surrounding humans; these plans are then taken into account when the robot plans its own tasks, satisfying the constraints imposed by the presence of other actors. Another example is [Alili et al., 2009], where a Human-Aware Task Planner (HATP) is presented, which integrates social behaviour rules. Moreover, social awareness should be taken into account in task execution when the tasks are collaborative or the task execution depends on the commitment of the persons. For instance, in [Clodic et al., 2009], the authors present a supervision system called SHARY, which takes into account not only task execution control, but also communications and interaction with humans in the achievement of tasks In [Sisbot et al, 2007], the authors present a path planner that takes into account motion models of humans and also models of their preferences, needs, etc., which are then encoded as a set of social constraints. These social constraints are, for instance, prioritizing the robots being in the field of view of the persons, not approaching a person from behind, etc. An A* search is then used to obtain the paths, taking into account the costs associated to these constraints. This way, the paths found also encode social acceptance.

In [Henry et al., 2010], a path planner based on inverse reinforcement learning is presented; the objective is to perform human-like motions with the robots. As the planner is learned for exemplary trajectories involving interaction, it is also aware of typical social behaviours. Moreover, efficient navigation in crowded environments requires taking into account human interaction models. In [Trautman and Krause, 2010], the authors consider the “robot freezing problem”. This situation occurs when the robot gets stuck due to the complexity of the environment, because the path planner is not able to find feasible paths, although these paths may exist. Usually, a person tracker is able to provide the position of the persons, and models are used to predict their future positions, which are in turn used to plan the path for the robot. However, these models in general do not consider human interaction (for instance, that humans will also try to avoid the robot). The predictions are performed considering that humans move independently of the situation, but actually if interaction is modelled, more efficient paths can be found. Models of person motion are mixed with models of typical interaction by using Gaussian Processes, leading to more efficient paths.

Beyond the state of the art

In FROG, 6DOF SLAM and localization algorithms based on vision and laser range-finders will be analysed under the circumstances of populated outdoor scenarios. Moreover, FROG will mix the results with the augmented reality tools. FROG will explore the combination of these with new sensors such as ToF cameras.

Regarding navigation, the main objective is the development of socially acceptable, efficient path planning and execution in crowded scenarios, both for robot navigation and group of persons guiding. By socially acceptable we mean human-like kind of motions when the robot is wandering or moving between way-points and objectives: by efficient, minimizing the amount of time when the robot is stuck due to the complexity of the scenario. In order to do that, online path planning taking into account uncertainties (future human trajectories) and “social” constraints will be developed. Human motion and human interaction models will be extracted and used in order to obtain more efficient paths.

Performance/criteria indicators

The developments in localization and SLAM will be evaluated in terms of localization and mapping errors by using the performance criteria developed in the FP6 Rawseeds project (http://www.rawseeds.org), and other related initiatives (like the SLAM evaluation toolkit from University of Freiburg, at http://kaspar.informatik.uni-freiburg.de/~slamEvaluation/index.php). Furthermore, the project will also consider the analysis of additional criteria related to the augmented reality tools.

Regarding navigation, the performance of the algorithms will be analysed in experiments through the project. One of the indicators will be the social acceptability of the navigation. However, as there is no current general metric for this, the project will analyse potential criteria with which perform comparisons in this topic.

4.1.3. Person Detection and Pose Recovery

There has been an extensive amount of research in computer vision on detecting people and their movement, see surveys (Forsyth et al. 2005, Moeslund et al. 2006, Poppe 2007). Most work has dealt with 2D approaches without explicit shape models; they bypass a pose-recovery step altogether and describe human appearance and movement in terms of simple low-level 2D features from a region of interest. Pedestrian detection has been a major application focus in this category; see survey by (Enzweiler & Gavrila, 2009). A second approach uses explicit a-priori knowledge of how the human body appears in the 2D image.

A prominent example in this category is the pictorial structure model (Felzenswalb & Huttenlocher, 2005). A third approach aims to recover 3D body-pose over time by matching articulated 3D graphical models. After feature extraction following one of the above mentioned approaches, an action recognition step typically follows. This involves statistical pattern matching approaches for time-varying data, see surveys (Mitra & Acharya 2007, Turaga et al 2009, Poppe 2010).

Beyond the state of the art

Most previous work assumes stationary, controlled backgrounds, single persons in the scene, several overlapping cameras, or a combination thereof. FROG will improve on the state-of-the-art by dealing with dynamic outdoor environments, possibly containing many persons, significant amount of occlusion and illumination changes. FROG will build upon state-of-the-art approaches to pedestrian detection (Enzweiler & Gavrila 2009), which will provide the necessary initialization for person tracking, at larger distances to the robot. At closer distances, FROG will consider algorithms to detect occlusion boundaries in the available 3D sensor data (either from TOF camera or from stereo vision), which will allow us to reason about visibility and the resulting weighting of component-based (i.e. per body-part) detectors, improving detection performance. FROG seeks to enrich localization information with an estimate of person body facing direction, adapting the mixture-of-experts approach of (Enzweiler & Gavrila, 2010b). Where visibility- and application based (e.g. computational) criteria allow, FROG will also recover 3D human pose, e.g. adapting the MAP (maximum a posteriori) estimation approach from (Hofmann & Gavrila, 2009) to a single view. 3D information is advantageous, since it enables the use of viewpoint invariant features for the analysis of human conversational signals and improves the accuracy of FROG's augmented reality-based HMI. A computational model will be developed to control the switching between the various levels of representation detail (i.e. position, position and body facing direction, and full pose recovery).

Performance/criteria indicators

Person detection and pose recovery algorithms will be tested by comparing system response to ground truth obtained by auxiliary sensors and/or human labelling. In particular, person detection will be based on the evaluation methodology described in the benchmark study (Enzweiler & Gavrila 2009) i.e. specification of region of interest and positional tolerances, and measurement of sensitivity and precision at the frame- and trajectory-level. The quality of 3D pose recovery can be assessed based on the evaluation metrics described in (Hofmann & Gavrila, 2009); ground truth for 3D body pose could, for example, be obtained by human labelling of the body joints in various views of a wide baseline multi-camera set-up.

4.1.4. Emotion Detection

The human face is our preeminent means of communicating and understanding somebody’s affective state and intentions on the basis of the shown facial expression (Keltner & Ekman, 2000). Given the significant role of the face in our emotional and social lives, it is not surprising that the potential benefits from efforts to automate the analysis of facial signals are varied and numerous (Pantic & Bartlett, 2007; Zeng et al, 2009). As far as natural interfaces between humans and machines are concerned, facial expressions provide a way to communicate basic information about level of engagement, interest, puzzlement, and other emotions to the machine (Pantic et al., 2007). Where the user is looking (i.e. gaze and/or head-pose tracking) can be effectively used to inform the machine about the user's current focus of attention. Also, combining facial expression detection with facial expression interpretation in terms of labels such as "joyful", "curious" and "bored" could inform the machine on the type of the feedback/ change needed and it could be employed as a tool for monitoring human reactions during web-based lectures, automated tutoring sessions, or tourist guide sessions as envisioned in FROG.

Because of its practical importance and the theoretical interest of cognitive and medical scientists (Ekman et al., 1993; Cohn, 2006), machine analysis of facial expressions and facial affective behaviour has attracted the interest of many researchers. For exhaustive surveys of the related work, readers are referred to: Samal & Iyengar (1992) and Pantic & Rothkrantz (2000) for overviews of early works, Pantic & Bartlett (2007) for a survey of techniques for detecting facial muscle actions, and Zeng et al. (2009) and Gunes & Pantic (2010a) for surveys of audiovisual (facial and vocal) methods for affect recognition in terms of either discrete emotion categories such as happiness, anger, fatigue, etc., or affect dimensions such as valance, arousal, expectation, etc.

Most facial expressions analysers developed so far target human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions such as happiness and anger.

Automatic detection of the six basic emotions (happiness, sadness, anger, disgust, fear, and surprise) in posed, controlled displays can be done with reasonably high accuracy. Detecting these facial expressions in the less constrained environments of human-computer interaction has also been explored recently (e.g., Koelstra et al. 2010, Nicolaou et al., 2010, Gunes & Pantic 2010b). Whilst state-of-the-art machine analysis of facial expressions is fairly advanced, it does suffer from a number of limitations that need to be addressed if it is to be used with freely moving subjects in a real-world environment as is the case with the robot-based system for monitoring interest and engagement of multiple people interacting with the robot to be developed in FROG. In particular, published techniques are still unable to handle natural scenarios typified by incomplete information due to occlusions, large and sudden changes in head pose, and other temporal dynamics occurring in natural facial behaviour (Zeng et al. 2009).

Body movement and posture are also predictors of affective states but they have been largely neglected because of a lack of a commonly-accepted set of emotional descriptors. Yet, they are accurate predictors of human affect (Ambady & Rosenthal 1992). In fact, perception of emotion has been shown to be often biased toward the emotion expressed by the body when facial and body expressions are incongruent (Meeren et al. 2005), and this perception has been shown to be robust even when cartoon-like characters or point-light displays are used (Pollick et al 2001). Yet, as in facial studies, most studies have focused on acted basic emotions and stereotypical body movement (e.g., dance; for an overview of the state of the art, see Gunes et al. 2008 and Gunes & Pantic 2010a). Natural expressions are more subtle than basic and stereotypical expressions, and approaches that rely on acted and often exaggerated behaviours typically fail to generalise to the complexity of expressive behaviour found in real-world settings.

Finally, agreement between humans rating an affective behaviour is greater when multiple modalities are combined (Ambady & Rosenthal 1992), and the dynamics of human behaviour is crucial to their rating (Ekman & Rosenberg, 2005). When it comes to fusion of multi-sensorial signals, past research has shown that this problem needs to be approached as the general classifier fusion problem, where correlations between input data streams (visual, audio, biophysical, etc.) are modelled while the requirement of synchronisation of these streams is relaxed (Zeng et al. 2009).

Past research has also indicated that the prediction of the input in one data stream based on the input in other data streams may be a more robust and effective approach to multi-sensorial signal interpretation than is the case with standard multi-modal data fusion (Petridis et al 2010). However, in most of the published studies on multi-modal analysis of human affective behaviour, the input from each modality is modelled independently and combined only at the final stage, which implements classifier fusion (i.e. decision-level data fusion). Although some attempts to make use of the correlation between multiple data streams have been made, it remains unclear how to model the observed multi-modal data on multiple time scales and how to model temporal correlations within and between different modalities.

Beyond the state of the art

In contrast to the existing approaches to human behaviour sensing and interpretation, the FROG project aims to develop methods for fully automatic, prediction-based multi-modal (vision-based, including facial, head, and body gesture modalities) recognition of spontaneous displays of basic and non-basic affective states including joy, engagement, interest, and boredom, observed in real-world robot-based guide sessions.

Performance/criteria indicators

For the face and the head, to be observed with a high-resolution camera, multi-person face detection, face tracking, and head pose estimation, regardless of head pose, clutter, and variations in lighting conditions need to be solved. The problem is difficult due the fact that not only the observed persons move and may occlude each other, but that the observation camera moves as well and may jitter due to the movements of the robot. Robust, fast and effective image registration should be developed.

4.1.5. Engaging Robot Social Behaviours and Personality

In recent years there has been an increasing interest in developing museum guide robots, which have been seen to have advantages over conventional audio and PDA guide systems. Robots' physical embodiment allows them to communicate with visible actions such as gazing and pointing in addition to verbal actions. This verbal and gestural communication raises people's social expectations of these robots and this in turn can be used to address multiple visitors simultaneously through such visible actions (Kobayashi et al., 2010).

Specifically in outdoor settings, there has been research on outdoor service technology such as automated wheelchairs Prassler et al. (2001). This work focuses mainly on collision avoidance (i.e. Schulz et al., 2003). In the outdoor context there is also work on human responses to robots in public spaces (e.g. Sabanovic et al, 2006). Kotaro et al. (2007) for example conducted an experiment in a train station. They used robots as a communication medium, presenting information about the travel duration to Osaka. Weiss et al. (2010) investigated people's willingness to support a robot that asked directions from passers-by. Such scenarios for robots navigating in public spaces and interacting with naive users focus on issues such as the ideal proximity between the human and the robot (Walters et al., 2009; Patchierotti et al., 2006), and how to initiate the interaction with humans (Bergstrom et al., 2008, Satake et al, 2009). However, most of the interaction between the user and the robotic system in public places is singular (only once) and short-termed, as passers-by have other goals than interacting with the robot. In the case of a robot in a dedicated cultural heritage site, the robot would have either longer or more frequent interactions with visitors. Also, visitors can be expected to be more focused on the context of the site they are visiting and thus, the context of interacting with the robot.

Robot Personality and social behaviours

It is believed that if a robot has a compelling personality, people will be more willing to interact with it and to establish a relationship with it (Breazeal, 2002; Kiesler, Goetz, 2002). In designing a robot personality, there are many challenges to consider. For instance, a robot may change its personality according to the user it interacts with or the personality could support a specific way of interacting (for instance a calm, quiet robot may afford stroking, an extrovert active robot may afford arousal).

Robot personality is conveyed in many ways. Emotions are often used to portray stereotype personalities such as friendly or grumpy (Yoon et al, 2000). A robot's embodiment (e.g. size, shape, colour), its motion, and the manner in which it communicates (e.g., natural language) also contribute strongly (Severinsun-Eklund et al., 2003). Finally, the tasks a robot performs may also influence the way its personality is perceived (Fong et al., 2003). Gockley et al. (2006) found that people unfamiliar with a robot preferred interacting with a moody robot, probably because the display of emotions was a novelty. Frequent visitors on the other hand preferred interacting with the positive version of the robot, probably because they felt a sense of common ground when they saw the happy expression. Tapus and Matarić (2006), showed that when robot behaviours were consistent with human personality types along the extraversion – introversion dimension, participants responded better when interacting with robots whose designed 'personality' matched their own. It has been shown that users perceive personalities in robot behaviours and appearances (Walters et al, 2008; Butler and Agah 2001). The need for consistent personalities for virtual characters has been pointed out by Isbister and Nass (2000). In research to date, robots and virtual characters have often displayed emotions in their interactions with people, but the specific role of personality and congruent social behaviours has yet to be assessed. The most developed robotic emotional model that we are aware of is the TAME architecture by Moshkina et al. (2003), which considers the four affective categories of personality traits, as well as attitudes, moods, and emotions; however, this model has not yet been fully implemented or systematically evaluated.

The perception of personality and social behaviours are strongly intertwined. In previous work, entertainment robots have been programmed with social behaviours in order to offer engaging interaction (for instance Furby as described by Turkle, 2006; and Pleo as described by UGOBE, 2008). Service robots have been outfitted with social behaviours to smooth incidental interaction with robots, for instance a hospital delivery robot that encounters humans in the corridors (Siino and Hinds, 2004). Therapeutic robots offer specific social behaviours to assist in therapy for instance for the elderly or for autistic children (Paro as described by Wada et al., 2005, Dautenhahn, 1999). To an increasing extent, research has been investigating the sustainment of long-term human-robot personal relationships with social behaviours (Freidman et al., 2003; Lee et al., 2009; Gockley at el, 2005; Bickmore and Pickard, 2005; Kahn et al., 2010, Breazeal, 2000, 2003a, 2003b). As described by Bickmore et al. (2008), relational agents must have a repertoire of behaviour that can be used to increase bonding with users. We will extend this state of the art by exploring social behaviours that are specific to guides and part of the strategies they use to engage their public (for instance focusing on a detail of a statue and pointing out an oddity for people to comment upon before explaining). Previous work on museum robots shows that robot personalities indeed add value to the visitor's experience (Sidner et al., 2004, Kuno et al., 2007, Yamazaki et al., 2009). Thrun et al. designed a museum guide to express happiness through its facial expressions when more visitors approached the robot (2000). Shiomi et al. (2007) also found that a robot can increase user engagement in a museum by referring to visitors by their names. As found by Bickmore et al. (2008), none of the agents described above use explicit models of the user-agent relationship, and they have a very limited repertoire of relational behaviour. Some are able to identify visitors (Shiomi, based on RFID tags, and Gockley, based on magnetic strip ID cards), but they only use this information to address users by name and both methods are limited in an outdoor setting.

Beyond the state of the art

In this project, we will extend the state-of-the-art in social robotics by identifying those synthetic personalities and social robot guide behaviours that will significantly impact user engagement and experience. We will extend the state-of-the-art behavioural science related to robotics by developing explicit models of user-robot engagement and an extensive repertoire of effective personality coloured guide behaviours.

Performance/criteria indicators

Progress is measured through laboratory and real-world user testing. Criteria indicators will include variables such as the extent to which users are engaged by the robot, the quality of the experience when interacting with the robot and user's emotional responses to the robot.

4.1.6. Multi-Modal Human Robot Interaction and Location-Based Content

Mobile robotic guide systems can make use of their physicality and location-awareness to adopt new techniques of entertainment and education of users. In the FROG project, robots will provide users with location specific information, specifically Augmented Reality (AR) content. Location based content and multi-modal interaction for FROG means that the robot will respond to non-verbal behaviours from small groups of users and present real images annotated with virtual contents accurately superimposed. For instance, the FROG robot may require pointing to highlighted objects on its screen ("Who wants to see what these structures looked like 1000 years ago?) and decide upon what AR overlay to provide depending on the feedback it receives from the group. Especially, gestures seem to lend themselves well for this type of interaction, allowing visitors to still communicate with each other. This is the first project, that adopts group gestural behaviours to interact with an autonomous social robot and the location based content it provides. Group behaviours such as gestures, touch and voice have been adopted previously in the context of ambient public displays or interactive tabletops such as in the work by (Morris et al., 2006 and 2004; Vogel and Balakrishnan, 2004; Yu et al., 2008; Kim et al., 2010). The multi-user games developed by Audience Entertainment – a US company partially owned by YDreams – are an example of robust software applications for large audiences that use gesture detection based interaction from multiple users.

Bennewitz et al., (2005) have explored interaction with multiple users with the goal to involve multiple persons into interactions and not to focus on a single person. They did so by detecting speakers as well as the robot displaying facial expressions and gestures. Even though they did find that people thought the robot was aware of them, the technological robustness of the robot was limited and its actions were not in the context of a specific location. In Human Robot Interaction, gestures have been mostly used in research that focused on interaction between one robot and one user (Waldherr et al., 2000; Kortenkamp et al, 1996; Riek et al., 2010), where gestures have included hand and facial movements and movement-based signaling of intent (Brazeal, 2003, Dautenhahn et al., 2006). As yet, no work has been carried out where a robot responds to location-specific user gestures.

The use of AR techniques has an additional benefit in the context of a mobile robot guide, as they can revolutionize the way people interact with unfamiliar environments in edutainment activities. Augmented Reality improves the user experience by enriching the presentation with context aware content. By tracking the user's position and orientation, complicated spatial information can be registered against the real world (Azuma, 1997) and the user can easily explore it through multimodal interfaces.

Beyond the state of the art

Mobile AR is a popular segment in the mobile apps market as shown by Layar and Wikitude – two AR mobile browsers. FROG will be, at least to our knowledge, the first robot that exploits mobile AR applications. This approach is new and requires further research in the fields of user interaction.

Till now AR applications have been mediated only by the display device, while in FROG AR applications will be mediated through a display device plus a socially aware robot.

Performance/criteria indicators

Performance of the AR component is measured with criteria such as the update rate for generating the augmented image and accuracy of the registration of the real and virtual image as well as in-body AR (a display screen embedded in the robot's body versus projected AR. Furthermore, the impact of the AR application and multi-modal interaction with it will be measured through user testing in laboratory and real-world contexts. Criteria indicators will include variables such as the level of user control, the extent to which users are engaged by the AR application and the quality of the experience when interacting with the AR application.

4.2. System Specifications

An initial division into subsystems was already presented at the DoW (see Figure 11). Based on the specified scenario (environment and envisioned use cases) this section describes the main specifications in each of the involved fields of expertise.

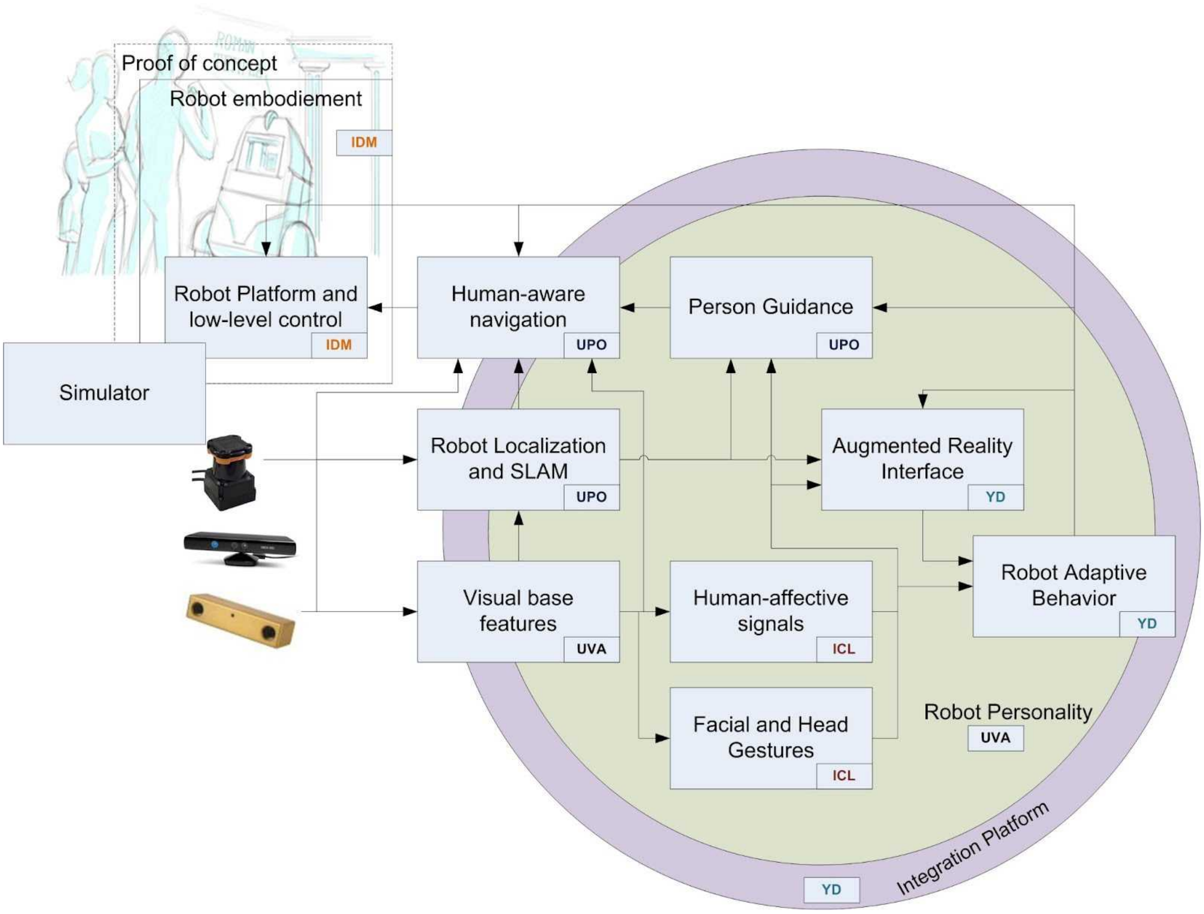

Figure 11. Overall technical strategy according to the roles of the partners

4.2.1. Robot Platform Robust Operation Specifications

To achieve the robust operation of the robot, the low-level architecture must include a group of low-level sensors associated with the robot basic functions (energy, locomotion, self-preservation, safety, etc.) that are ruled by dedicated low-level control loops. These low-level loops must react to changes that can affect the robot operation and also provide fault detection information to the high-level behaviour trees.

The high-level behaviour trees must adapt the robot operation to deal with changes in the information related with the environment but also with the changes on the low-level dealing with fault detection information.

The low-level hardware must be able to run self-diagnostics for each board, have a fault detector for each component and provide self preservation procedures for the system.

Self-Diagnostic (of each electronic board)

The hardware architecture must allow the running of low-level self diagnostic control loops. The control loops must systematically check the robot hardware/software and communication variables to detect any problem that may jeopardize the normal function of the robot.

Fault Detection (of each main component)

Each main component should have fault detection hardware attached to allow the detection of any system operation failure and to report it.

Self Preservation (of the robot)

Two types of sensors are going to be installed for self preservation procedures.

- **Environmental sensors** (Self preservation)
  The robot must be aware of weather condition changes that might affect its performance. For that reason the following sensors must be integrated on the robot:
    - **Rain** - This sensor detects the presence of rain or presence of water that can put some of the robot components at risk;
    - **Temperature** - This sensor will give the information about the ambient temperature;
    - **Humidity** - This sensor will give the information about the relative humidity of the environment.
- **Robot sensors** (Self preservation)
  The robot must carry sensors that allows it to monitor the system and take adequate measures in case of critical events:
    - **Battery level** - Detect the battery level and estimate the remaining time of operation.
    - **Motor temperature** - Measure the temperature of each motor and determine whether it is running within the manufacture specifications.
    - **Driver temperature** - To measure the temperature of each driver and determine whether it is running within the specifications.
    - **Bumpers** - to detect collisions and take evasive manoeuvres (also to provide human safety).
    - **Sonars** - to create a safe/danger area of operation (also provide human safety).

One important element for robot self-preservation is an external docking station where the robot is able to connect and charge the batteries. The docking station will be located in a remote service area where the robot can plug itself in and power the battery charger that it has on-board.

The FROG robot must be designed according to industrial standards, with an adequate robot chassis and locomotion to deal with the versatile terrain, with a high degree of resistance to unfavourable environmental conditions (e.g., dust and water).

4.2.2. Robot Navigation and Localization Specifications

The robot localization component should provide an estimate of the localization of the robot in 6DoF. 3D orientation is required for the augmented reality devices, which will be used to project information on the screen or using some projectors, as indicated by the scenario description. Moreover, these scenarios are inherently 3D. A map-based localization approach will be pursued, and therefore, a Simultaneous Localization and Mapping (SLAM) phase will be required during the robot deployment operation.

Localization and SLAM

An offline full-SLAM module will be implemented to build the maps required for localization, augmented reality and navigation.

Vision will be the main modality, although laser information will be used as well (and GPS if available, although in the scenarios considered the GPS coverage is very limited).

The module requires as inputs the sensorial information from Section 5.1:

- odometry and other navigation sensors (Section 5.1)
- laser information
- stereo pairs

The map should contain also relevant information for the augmented reality content. Therefore, it should be augmented with features and locations related to the scenarios defined.

Localization

The main inputs to the localization module will be the sensorial information, namely:

- odometry and other navigation sensors (Section 5.1)
- laser information
- stereo pairs
- visual features from the visual base features module
- SLAM: map built in the SLAM phase

The pose estimation will be provided to the rest of the modules of the system.

The target values for both sub-modules will depend on the particular scenario, but for navigation and augmented reality purposes a localization mean error below 50 cm and orientation mean error of the order of 2-3 degrees will be pursued.

Human-aware navigation

The module should provide the navigation stack to cope with the scenarios depicted in Section 2. Two main modalities can be identified: way-point navigation (including people approaching) and people guidance. The latter is considered by the Person Guidance subsystem, although it is very related to this module.

This module interfaces with other modules.

The navigation component will receive information from:

- Localization: 6DoF pose of the robot
- Visual base features: location and pose of persons within the field of view of the cameras
- Robot adaptive behaviour: high level commands and way-points

The component will output signals to the low-level control in the form of velocity commands.

The metric will be the percentage of navigation total failures (that is, failures in navigation tasks that require an operator intervention to recover) during the time of operation (week). As the robot is intended to operate 24/7, the target value is 0. Also, the navigation has to take into account social constraints and the presence of humans in its performance.

Person Guidance

The person guidance module is required to cope with the basic scenario. It can be seen as a different modality of the navigation module. In this case, persons and robot have a common goal, the next way-point to reach. Therefore, the robot has to adapt its motion to the motion of the persons.

Besides the inputs described before for the navigation component, this modality will also receive:

- Visual base features: head tracking results
- Human-affective signals: level of engagement of persons

According to the scenario, the module should be able to guide the persons to a certain waypoint, while adapting its pace.

4.2.3. Perception Specifications

Person Detection and Pose Recovery

The person detection and pose recovery is required to cope with the basic scenario. Person detection acts as the initial trigger for the FROG robot to approach visitors at distances up to 25m. Pose recovery will be used to estimate whether persons are engaged with each other or with a nearby exhibit, at distances of up to 15m.

Input to the person detection and pose recovery module, at each time step:

- Stereo image pairs (stereo system is assumed to be calibrated off-line)
- Any constraints on accessible visitor walking areas in front of the robot, as derived from the localization module and digital map (i.e. obtaining "regions of interest" where to search for persons in the images).

Output of the person detection and pose recovery module, at each time step:

- $\{(ID1, X1, Z1, \Theta1), ..., (IDN, XN, ZN, \Theta N)\}$ where ID is a unique person identifier, and X, Z are the lateral and longitudinal positions, respectively, Θ is a representation of the person pose, all with respect to a robot-centred coordinate system, for each of the N persons detected. In the basic scenario, Θ involves a single value Θ, the overall person body facing direction. In the advanced scenarios, Θ could for example also identify head and/or arm poses.

Performance metrics will involve the number of correct vs false person detections, the number of ID switches, localization and pose estimation accuracy.

Emotion Detection

The main aim of the emotion detection component is to detect engagement and boredom of the visitors engaged with the FROG robot. Given that the robot should operate in outdoor, possibly crowded environments, the emotion detection module should be able to cope with face tracking, head pose estimation and facial expression analysis regardless of head motions, clutter, and variations in lighting conditions.

The main input to the module should be the high-resolution-camera feed that will capture the faces of people that are close to the robot, interacting with it. The camera should be placed in the following way:

- 1.5m from floor, in the centre of the robot, with the possibility to modify tilt
- the camera mounting should be with the base plate underneath and it has to be mounted on a 1/4" camera screw, and
- the camera casing needs to have a hood to prevent lens flare from sun shining on the lens.

To enable video data processing, a laptop dedicated to the emotion detection component should be mounted on the robot. The estimated weight of the laptop is up to 5kg. The laptop casing needs to be air-cooled, with extra space for usb/ethernet connectors on back and sides of the laptop.

To potentially enable detection of who is speaking when, a Kinect camera containing a microphone array should be mounted on the robot as well. The camera should be placed under or above the screen, in the open air (possibly under a small roof).

4.2.4. Interaction and Content Exhibition Specifications

The end user workshop provided a starting point to study the possible adequate content to include in the guided tour. The visitor experience documentation, the user requirements and the human tour guide analysis, presented in this deliverable, will provide the fundamental baseline to design and develop the content presentation and interaction strategies.

Empirical conclusions so far, lead us to assume that the portfolio of interaction technologies and content presentation proposed in the DoW are adequate for the test sites and will cope with the objectives.

In terms of the presentation components, the strategies and conclusions are as follow:

Presentation components

- LCD Display on the robot body
  - This strategy has been used in many previous robotic projects mainly with two objectives; to present multimedia content or has a versatile way to show evidences of affective robot states, for instance through a simulated virtual robot face. In the case of FROG, a transflective display with an adequate dimension, will be an important vehicle to present information. The limitations in terms of visibility for a group of people is also a reality in the observed guided tours, where the human guides show a photo they carry on their information pack. In these situations the guide shows the photo around the group with each visitor leaning in individually, a solution that the FROG robot can effectively mimic.
- LED video projector
  - The capabilities of actual off-the-shelf pico-projectors and the popularity of video mapping (also referred to as spatial augmented reality) will represent an attractive way to present content. FROG will be able to directly superimpose information through its video projector directly onto the scenario, augmenting the vision or revealing, otherwise, covered information. Although this solution presents several constraints in terms of direct sunlight, the two test cases present several permanently shaded areas where it can be effective. The limitation in terms of projection power and light conditions in outdoors will not allow for large projections but the scenarios present many opportunities for short range projections on objects.
- Laser pointer
  - The study performed on location made evident the importance of a guide pointing to the object he is presenting. The safe laser pointer in a pan/tilt arm mentioned in the DoW proves to be a very effective way to achieve this with much higher precision as, for instance, would be important in the case of Alcazar´s richly ornamented ceilings.
- Audio
  - Audio plays a very important role in any guided tour, both for content presentation and for visitors' orientation and guidance. It is also a very important aspect of personality and affective expression. The FROG robot will include an audio component that will encompass all these aspects. It is important to strengthen the fact the this project will not implement a conversational agent.

Interaction modalities

User interaction in FROG will be mainly supported by computer vision. Although a touch-screen will also allow for contextual input from the users and audio input will be used, mainly for sound source localization, the FROG robot will rely strongly on its various vision systems to provide multi-user interactivity.

- Implicit interaction
  - The perception and emotion detection components will allow for the adaptability and adequation of content presentation strategies and tour rhythm, based on users attention and level of engagement.
- explicit interaction
  - Gesture analysis will provide natural user interfaces for explicit interaction with the users.

5. Robot Platform Hardware Specifications

Based on the scenario analysis, on the system requirements and specifications, the robot platform has been specified to meet the following overall features:

Robot Platform Main Features

- Robot Kinematics: 4 wheeled differential drive (pneumatic tires for uneven terrain)
- Predicted Weight: 50 Kg
- Payload Capacity: 30 Kg
- Predicted Height: 1.4 to 1.5 m
- Battery Autonomy: 4 to 6 hours
- Maximum Velocity: 1.5 m/s
- Low-level sensors: battery level, motor encoders, bumpers, ground sensor, sonars, temperature, humidity, rain
- High-level sensors: lasers, IMU with GPS, standard cameras, depth cameras, stereo cameras
- Actuators: 2 DC motors for locomotion
- Interaction devices: monitor for AR contents, sound, microphone, led lights.
- Other interaction devices such as laser pointers and image projectors are also being considered.

In terms of kinematics the platform will be running over a four wheeled differential drive, where in each side of the robot two wheels are connected to a single motor. This type of kinematics is adequate for an uneven outdoor (or indoor) terrain, and the robot will deal easily with small obstacles. We have also considered the use of a two wheel differential drive with a caster wheel, but with this type of kinematics the robot could face some difficulties to overcome small obstacles in contact with the caster wheel that has no traction.

The predicted weight of the robot has been estimated by calculating the weight of batteries, motors, computers, inverters, internal structure, outer shell and all the add-on equipment that is being selected for the project. The predicted height of the robot has been estimated by the placement of the perception cameras specified by the partners.

The following sections describe the specified main components of the platform.

5.1. Sensors

The robot will be equipped with perception, navigation, interaction, environment and low-level safety sensors. For **perception** the robot will be equipped with cameras, providing a middle range perception (<25m) and short range perception (1 to 2m). For **navigation** the robot will make use of encoders to control the velocity of the motors, an inertial sensor with GPS and a stereo camera to determine the position/orientation in the environment, lasers to detect obstacles and mapping, and standalone camera for guidance missions. For **interaction** the robot will use standalone cameras for people tracking, face analysis and body gestures, and also microphones. For **environmental** sensing the robot will be equipped with temperature, humidity and rain sensors. Finally, the bumpers and sonar sensors will provide **low-level safety** sensing.

Here follows a list of sensors that have already been selected to be used on-board.

5.1.1. Perception Sensors

The robot will make use of different cameras for navigation in the environment, features identification, pedestrian detection and tracking, body orientation estimation, face and gesture analysis. It can also be used to detect changes in the surrounding environment.

- Front Stereo Vision Camera: Dalsa Genie-HM1400 XDR (2x)
    - Function: localization, obstacle detection, pedestrian detection and body orientation estimation;
    - Position on Robot Platform: looking ahead; 1.2 m high
- Front Vision Camera: Dalsa Genie-HM1400 XDR (1x)
    - Function: face analysis and fine body gestures.
    - Position on Robot Platform: looking ahead; 1.2 to 1.4 m high
- Rear Vision Camera: Dalsa Genie-HM1400 XDR (1x)
    - Function: people tracking for guidance mission
    - Position on Robot Platform: looking back; 1.2 m high
- Front RGB Camera
    - Function: Augmented Reality
    - Position on Robot Platform: opposite direction to the display

5.1.2. Navigation Sensors

The robot will navigate in the environment while fusioning the measures provided by different sensors. Outdoors the robot will be able to use the stereo pair, lasers, GPS (where available), and the odometry and the inertial sensor to estimate its posture. For the robot obstacle avoidance and mapping it will be using the lasers and sonar sensors. The front stereo vision camera described in 5.1.1. will also help in the navigation by detecting obstacles and persons.

- Inertial Sensor IMU with GPS: Xsense MTI-G
    - Function: Localization of the robot (position and orientation)
    - Position on Robot Platform: as close as possible to the robot´s centre of gravity
- Front 2D laser range-finder: Hokuyo’s UTM-30LX
    - Function: localization and obstacle avoidance
    - Position on Robot Platform: frontal and horizontal
- Front 2D laser range-finder: Hokuyo’s UTM-30LX or URG-04LX
    - Function: obstacle avoidance (3D perception of ramps, slopes, etc)
    - Position on Robot Platform: frontal, tilted or vertical orientation
- Rear 2D laser range-finder: Hokuyo’s UTM-30LX
    - Function: obstacle avoidance
    - Position on Robot Platform: on the rear and horizontal
- Sonar Sensors: XL- MaxSonar®- WRC1™
    - Function: obstacle detection (ex:glass wall or objects)
    - Position on Robot Platform: ring of sonars around the robot

5.1.3. Interaction Sensors

The front vision camera described in 5.1.1. will be also used to sense visual user feedback for natural user interaction. There will be also the following sensors:

- Microphone array: Microsoft
  - Function: user sound feedback for natural user interaction
  - Position on Robot Platform: turned to the users
- Touch-screen
  - Function: user feedback on specific contents
  - Position on Robot Platform: turned to the user

5.1.4. Environment Sensors

The environment sensors will be used to detect variations on the weather conditions that can affect the normal operation of the robot. These sensors are: temperature sensor; humidity sensor; and rain sensor.

5.1.5. Low-level Safety Sensors

The fundamental sensors for low-level safety will be the sonar sensors described in 5.1.2. and the bumpers switches.

5.2. Actuators

The robot will be equipped with locomotion and “interaction” actuators.

For locomotion the robot will make use of two motors to actuate the 4 wheels of the robot. Each motor will actuate one pair of wheels in each side of the robot (differential drive).

For interaction there will be one touch-screen to display contents (e.g. AR contents), stereo speakers and led light colouring specific parts of the shell. Other types of interaction devices will also be evaluated: laser pointer and image projector.

5.2.1. Locomotion Motors

With the specification of the robot it was possible to calculate the required power to drive the robot. It has been specified a robot weight of 50 Kg with an additional payload of 30 Kg and running at a maximum speed of 1.5m/s. To calculate the required motor and gearbox torque we have defined three slope inclinations that the robot should be able to overcome: 5, 15 and 25 degrees. For the motor calculation a standard wheel with a radius of 0.155 m has been considered.

Description	5 degrees	15 degrees	25 degrees
Robot Mass	50	50	50
Robot Payload	30	30	30
Wheel Radius (m)	0.155	0.155	0.155
Gravity acceleration (m/s2)	9.8	9.8	9.8
Robot Force (N) for X degrees	68.33	202.91	331.33
Torque @ Wheel Axis (Nm)	10.59	31.45	51.36

Once the required wheel axis torque has been calculated, the gearbox and the gear connection between the wheel axis and the gearbox axis that is able to support the calculated torque will be defined. The wheel will be connected to the gearbox through a set of timing pulleys with a reduction of 2.45:1. The chosen gearbox is a Maxon Gearmotor with Planetary Gearhead GP62 19:1.

Description	5 degrees	15 degrees	25 degrees
Torque @ Wheel Axe (Nm)	**10.59**	**31.45**	**51.36**
Timing pulley set reduction	**2.45**	**12.84**	**2.45**
Torque @ Gearbox (Nm)	**4.32**	**0.155**	**20.96**
Max. Continuous Torque(Nm)	**25**	**25**	**25**
Intermittently permissible torque at gear output (Nm)	**37**	**37**	**37**
Max. Efficiency (%)	**75**	**75**	**75**
Torque @ Motor Axe (Nm)	**0.3**	**0.9**	**1.47**

The gearbox will connect to a motor that is able to provide the required torque to drive the robot. The chosen motors are the Maxon RE50 of 200W at 24V.

Assigned power rating: **200 W**
Nominal Voltage: **24 V**
No load speed: **5950 rpm**
Stall Torque: **8920 mNm**
Max. continuous torque: **405 mNm**
Max. Velocity: **1.5 m/s**

From the data calculations it is possible to conclude that the motor can be running continuously at maximum speed when the slope inclination is less than 7 degrees. Considering slopes with higher inclination it will be able to provide the required power for short periods of time.

Motors	Part Number	Quantity
Motors	**Maxon RE 50 200W 24V**	**2**
Gearbox	**Maxon GP 62A 19:1**	**2**
Encoder	**HEDS 5540 - 500 pulses**	**2**

5.2.2. Interaction Actuators

Here follows the list of interaction devices. The robot will be able to display the contents in an onboard monitor or to project them over a surface. Only one of the listed projectors will be included.

- Display with touch-screen
  - Function:content display (e.g., AR contents)
  - Position on Robot Platform: turned to the user
- LED projector: 3M MP180 or Optoma Pico PK301 or AAXA P2 pico
  or

Laser projector: Microvision SHOWWX+ Laser Pico Projector or AAXA L1 v2 Laser Projector

  - Function: projection on POIs (Spatial augmented reality)
  - Position on Robot Platform: should be mounted over a small pan&tilt unit

- Stereo Speakers
  - Function: sound exhibition of contents; robot communication
  - Position on Robot Platform: turned to the user

5.3. Electronic Power Architecture

The robot will be powered by two pairs of 12 V batteries. One of the pairs will provide power to connect the computers and all the electronics. The other pair will be used to provide energy for the motors. A charging unit will be developed and used inside the robot to charge the batteries. The batteries and the power in the robot will be managed by a sensor and a charger board that will measure the battery level, battery charge, and also control the units (motors, sensors, actuators and inverters) powered by the batteries.

All the on-board electronic systems will be powered by the battery system. The ATX computer power will be providing regulated voltages (from 5 V to 12 V) and it will be also included an inverter able to provide 600 W 230 VAC to any device requiring AC power supply (e.g. monitors). Figure 12 depicts the on-board power architecture.

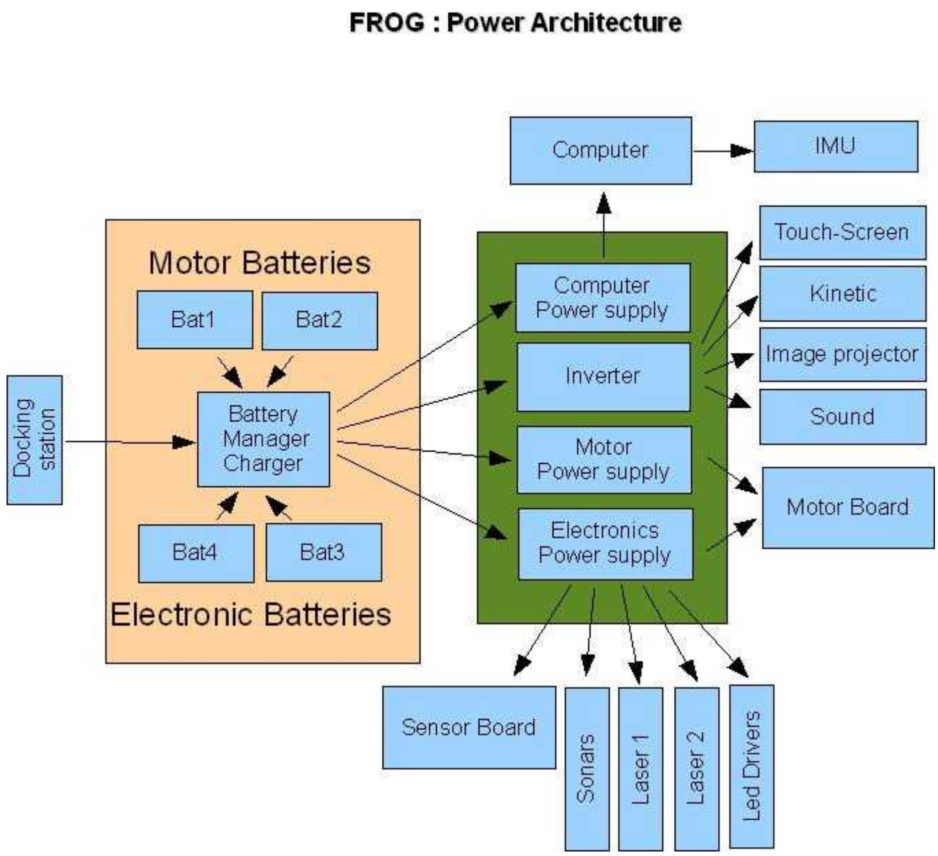

Figure 12. Onboard Power Architecture

5.4. Low-level Communication Architecture

The on-board robot navigation computer will communicate with the two boards (Sensor & Charger Board and the Motor Board controller) using 2 USB ports. In each board there will be a USB to RS232 converter that will convert the USB data packages to serial RS232 packages for the Boards controllers.

Each Board controller will communicate with the other allowing the exchange of information in-between them. This communication channel will allow the execution of low-level behaviours, for example, react against an imminent collision, enter in charging mode with motors shut-down, reduce the motors velocity when the batteries are low or react to changes that can affect the robot's operation, which is fundamental to the improvement of the overall system dependability.

The main controller from the Sensor&Charger Board will communicate with other microcontrollers using I2C communication ports. The main controller will act as the master and the other microcontrollers will behave like slaves. The Sensor&Charger Board will connect to the battery charger, sensor, led lights and sonar acquisition boards. The Motor Board controller will connect to the PI Motor controllers and also to temperature sensors.

Each controller will have a low-level fault diagnosis that will check the operation state of each microcontroller and also monitor all the communication in-between the devices. The low-level communication architecture is depicted in Figure 13.

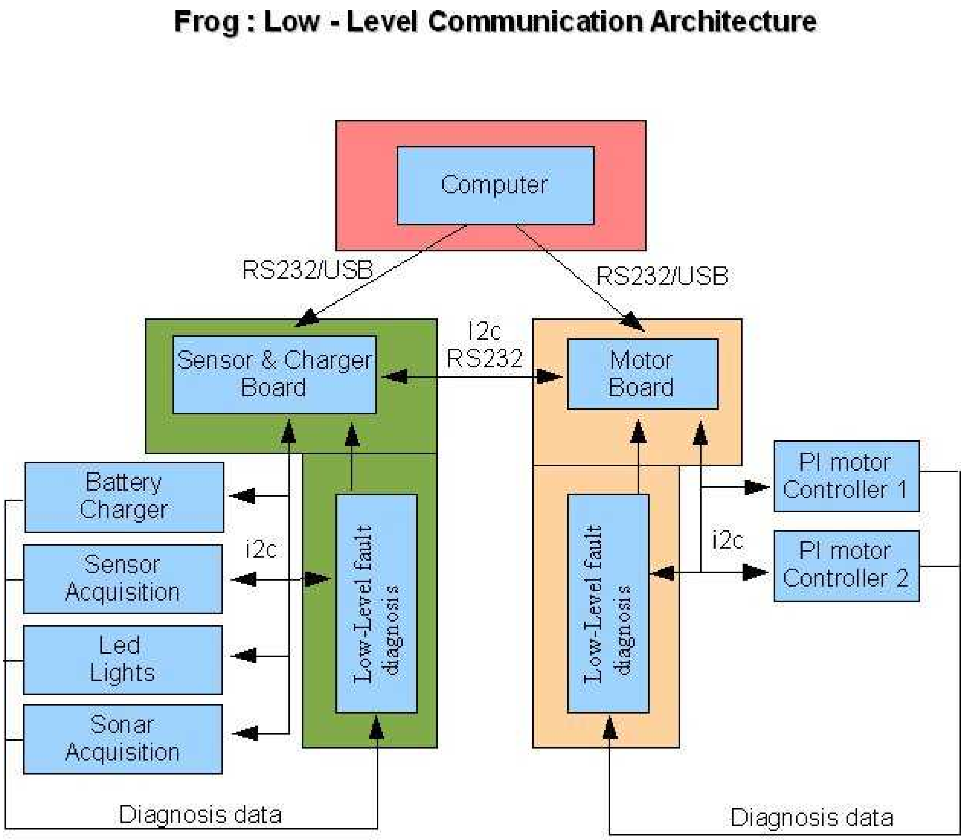

Figure 13. Low-level Communication Architecture

5.5. Motor Controller Board

The Motor Controller Board will manage the robot locomotion. It will receive orders from the high-level robot navigation computer and return the information from the encoders.

The Controller will connect to two PI Microcontrollers that will generate the control actuations to follow the velocity references. Each microcontroller will connect to the motor using a 1000W H-bridge, and will measure the current and temperature of the motor and provide the pulses measured by the encoder.

Each microcontroller will be optically isolated from the motor driver using a high-speed optocoupler for the control actuation signals and an optical amplifier for the current measurements. It will also be optically isolated from the computer communication port, again using a high-speed bidirectional optocoupler. Several low-level fault diagnostics will be implemented to detect problems in the normal work of each component and lacks of communication. The Motor Board will communicate with the Sensor&Charger Board exchanging information about the system status, diagnostics and environment condition, allowing low-level robot reactions to changes that can affect the robot's operation. Figure 14 depicts the architecture of the Motor Controller Board.

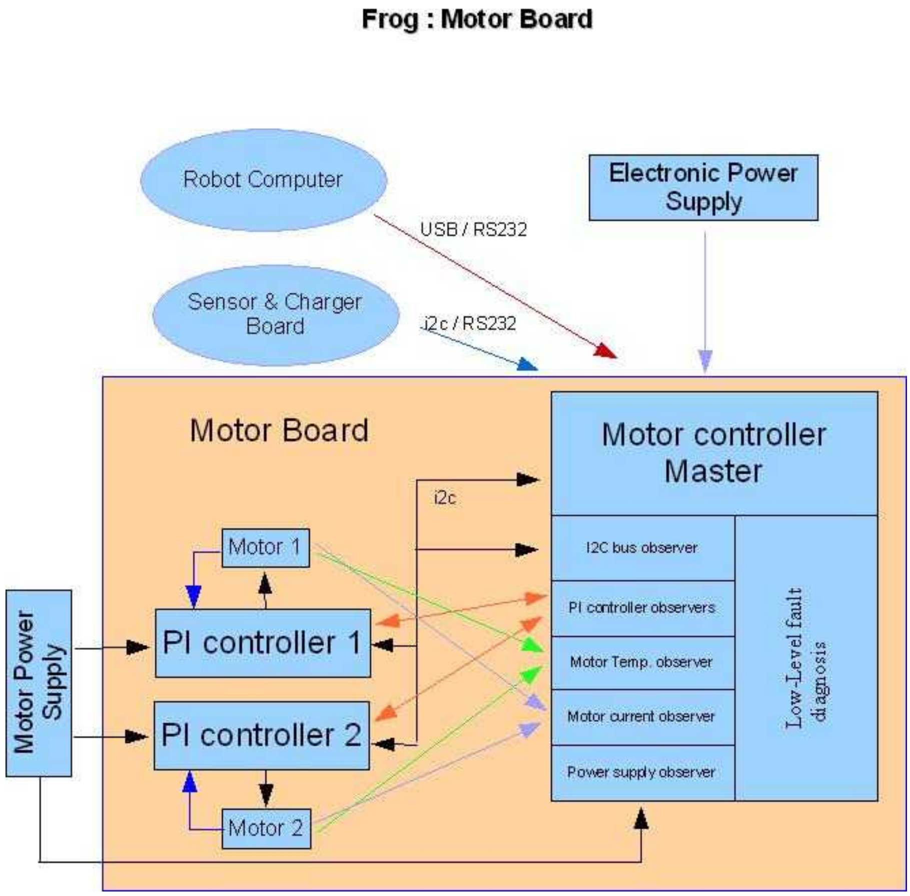

Figure 14. Motor Controller Board

5.6. Sensor and Charger Board

The Sensor&Charger Board (see Figure 15) will be responsible for the power management and also the sensor acquisition. It will receive orders from the on-board robot navigation computer and return information about the batteries, sensors and actuators.

It will manage all the power system, measuring the level of energy in each battery, connect or disconnect the power of the devices, manage the connection to the Charge Docking station and control the charge of each battery. It will also be responsible for connecting a set of sensors and actuators that will be used in the project. Several low-level fault diagnostics will be implemented to detect problems in the normal work of each component and communication. The Sensor&Charger Controller will analyse the gathered information from the sensors and will run low-level control loops that will check for critical changes in the environment or system that can affect the robot operation.

The Controller will have one dedicated channel to communicate with the Motor Controller, allowing it to send direct (and fast) commands to the motors and also getting information from them.

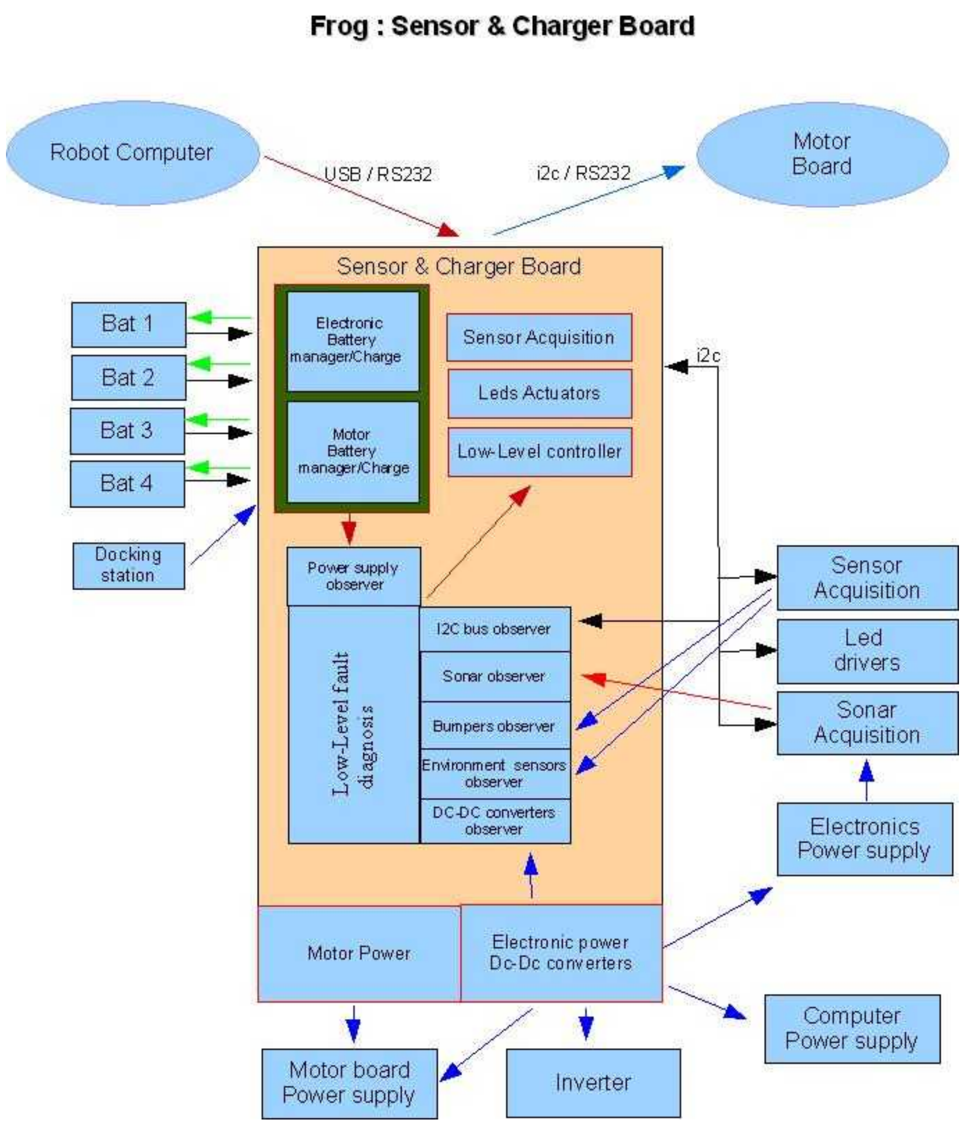

Figure 15. Sensor and Charger Board Architecture

5.6.1. Charger Docking station

One important capability of the robot is the possibility to work without human intervention. To achieve this point the robot must be able to manage the on-board power and autonomously charge itself. One charger docking station will be developed and installed in a service area, where the robot can enter and plug itself in. This docking station will provide the necessary power that the on-board battery chargers need to charge the batteries. The docking station is a passive power station and all the control of the charging process is managed by the on-board Sensor&Charger Board.

5.7. Robot Platform Design (First Thoughts)

To determine the kinematic problems, to have an idea of volumetry and weight with the integration of some common components, some first drawings of the platform have been developed. A robot base platform has been designed taking into consideration two wheel diameters: 300 mm and 250 mm. The platform (see Figure 16) consists of a four wheel differential drive base with each motor connecting 2 wheels on each side, four 12 V 20Ah lead-acid batteries, computer system for platform control, Motor Control Board, Sensor&Charger Control Board and Inverter.

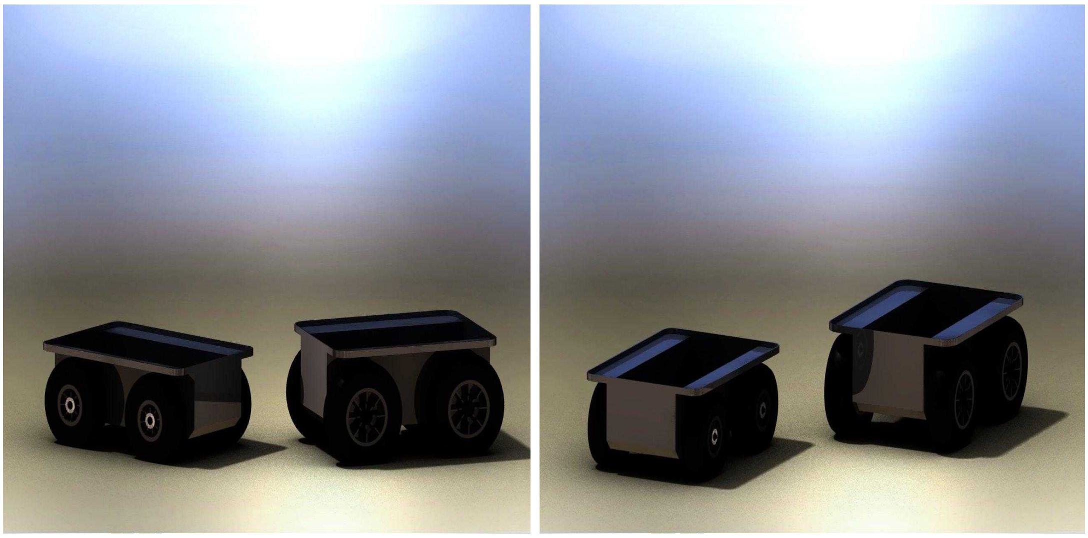

Figure 16. Four wheel differential drive platform

Based on this design it was possible to determine the centre-of-mass of the robot and have a first idea of the dimensions (and available space) of the different components.

Based on this design of the platform, a simplified outer-shell has been designed to have an idea of dimensions in comparison with a person (Figures 17 and 18).

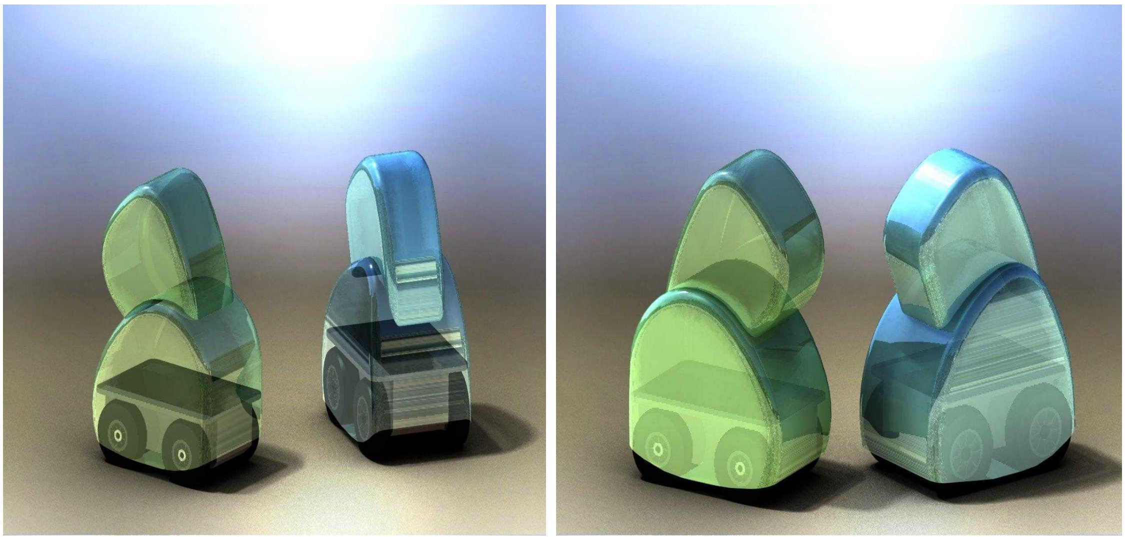

Figure 17. Simplified outer shell

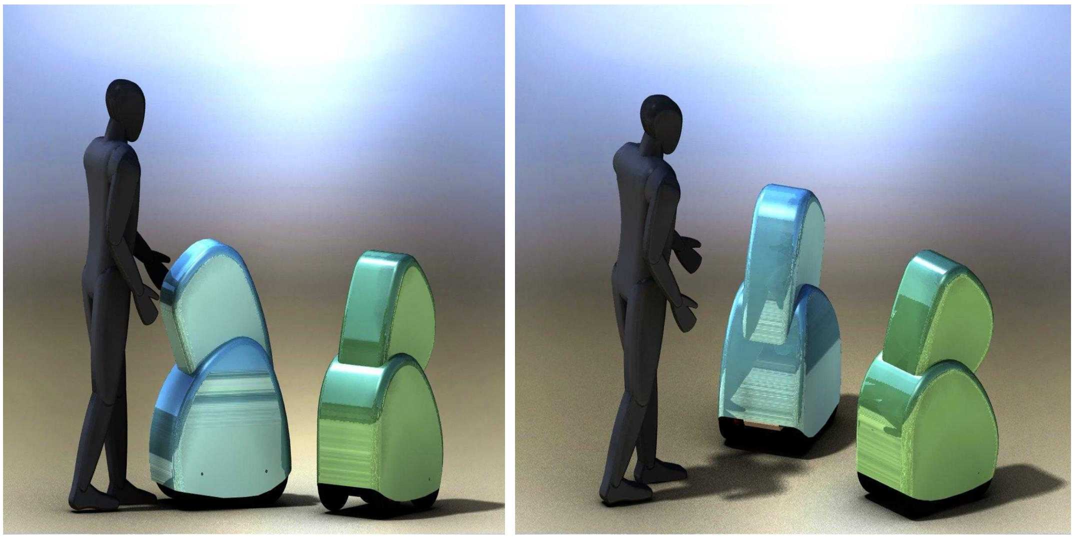

Figure 18. Comparison with a person

5.8. Data Collection Robot Platform

The proposal includes an iterative approach to data collection. There will be a workshop (month 8) to collect data from the selected environments, using an existing robot platform equipped with the project perception and navigation sensors. The platform will run outdoors (not on a rough terrain), have wireless control with an external laptop (plus joystick), and observe a small number of people and register data with its sensors. This data collection will be an important input for the various instantiations of the actual system being built.

For this data collection workshop the consortium will make use of an existing IDMind robot platform. This is a two wheel differential robot with one caster wheel on the back. It uses two 150 W Maxon motors with two PI controllers with 250 W drivers for locomotion. For navigation it has an onboard computer with wireless communication. To power the system it uses two 12 V 20 Ah batteries to power the motors and electronics.

For the data collection the following equipment will be also integrated in the platform:

- One 12V 40Ah battery;
- One 600W Power Inverter;
- One desktop for **UvA**
- One laptop for **ICL**
- Two external hard drives for Lacie 2TB Minimus USB 3.0 Desktop Hard Drive for **ICL**
- One set of stereo cameras for **UvA** and **UPO**
- One stand alone camera for **ICL**
- One 22" touch-screen Monitor for **ICL**
- One Kinetic Xbox camera for **ICL**
- One laser Hokuyo UTM-30LX for **UPO**
- One inertial sensor XSens MTi-G for **UPO**

Figure 19 depicts the platform with all the equipment to be used in the data collection workshop.

Figure 19. Data Collection Robot Platform

6. Integration Architecture

The Integration Architecture for the FROG project is, at this point of the project and following its planning, a work in progress. A survey on possible architectural frameworks has been completed (D5.1) as the first step for the integration task. This survey compares several software frameworks with the objective of selecting the most suitable to be adopted in the FROG EU project.

The architecture of FROG can be conceptually divided into three different layers, the low-level control algorithms layer (comprising motor control, safety-stop, low-level sensors abstraction layer), the middle-level control algorithms (comprising the navigation and localization systems), and the high-level behaviour and interactive layer responsible for emotion detection and human-robot interaction via augmented reality tools.

The key requirements for the FROG framework were identified and imposed the study of the following software framework families: robotics frameworks, augmented reality frameworks and game frameworks.

The survey presents an analysis of some of the most widely adopted robotic frameworks, namely, Player, ROS, YARP, CARMEN, OROCOS, Orca, MRDS, URBI, MRPT, MOOS, OpenJAUS, ERSP, CLARAty, and GenoM.

Augmented reality denotes an online view of the real world, augmented by computer-generated information such as sound, video, and graphics. This technology results in an enhanced perspective of the world by the user. This is far different from virtual reality, which replaces the real world with a simulated one.

Total Immersion D'Fusion and Metaio Unifeye are augmented reality frameworks that provide the basic tools to develop augmented reality applications. Game engines also provide useful tools such as rendering of 3D images and physics simulation. The game engine Unity provides tools for creating 3D video games or other interactive content such as architectural visualizations or real-time 3D animations. The multimedia augmented reality framework YVision is also presented and its key features emphasized.

The basic tools provided by these frameworks will be exploited during the project to construct algorithms in the areas of localization and navigation, people and emotion detection, and augmented reality to create an appealing interaction with tourists.

It was concluded that the best way to address the project complexity and promote the reuse of the algorithms in other potential applications is by means of a modular architecture, where each module fulfils a specific task and where the modules communicate with each other through a middle-ware to exchange information. This best practice also couples nicely with the practicality of using the best tools for the job. As such we have analysed and discussed the advantages of two alternative solutions to be adopted in the FROG project, one based solely on YVision and other based on a two frameworks solution, ROS and YVision. Where ROS can also double as the middle-ware communication opening the opportunity to include other frameworks that have bindings or integrate to ROS.

In the latter the ROS framework is adopted as the base for the development of the low-level control layer and the middle-level control layer. On the other hand, ROS is not so suitable for multimedia application. Hence, the high-level behaviour and interactive layer have to be developed using an other framework namely YVision. This alternative requires the development of a communication module that enables the exchange of information between ROS and YVision. Some proof of concept should be developed in the near future.

The next steps to the Integration Architecture will be to specify what each of the modules requires from the others in terms of data information and also data each of them produces. Also we will be defining the way modules and tasks are controlled by the higher layer. In parallel we will define the machines where each module runs and which sensors are connected to which machine.

7. Methodology of Development

While the FROG project is addressing challenging problems, the given combination of partners guarantees substantial breakthroughs within the planned time frame. The research has been decomposed into four phases with accompanying milestones. The work plan and the derived work package structure will maximize the synergies within the consortium, which is a unique combination of complementary partners:

- **UvA** will be in charge of overall management, identification of the requirements in the context of outdoor guidance, body pose detection, and cognitive social psychological user studies on the effects of robot synthetic personalities and guide behaviours.
- SME partner **YD** will develop the adaptive human robot interaction dialogue system, the internal logic for the robot and the simulation environment to evaluate the project systems. They will also contribute to gesture and course audio recognition and develop the augmented reality content.
- SME partner **IDM** will design and construct the outdoor robot hardware and low-level systems for safety and robust outdoor operation
- **UPO** will develop SLAM and 6DOF based localization techniques and navigation algorithms in crowded outdoor settings
- **ICL** will contribute to sophisticated human behaviour and affective state recognition in populated outdoor settings.

The development of a research tool. In the realization that FROG cannot claim to deliver the ultimate socially competent tourist-guide robot in the three year time period, we instead plan to build a reliable research platform that allows the investigation, within and beyond the project duration, of the kinds of phenomena that occur in interactive guide-related conversations between a human and a robot, such as:

- How can users be engaged in viewing the target tourist attraction? How can the users' interest level be increased?
- How do users react to robot responses that are personality-coloured? Can this be used to increase the users' engagement?
- What are the minimum robot responses and what is the speed of providing these responses for the interaction to be acceptable?

Again, we cannot hope to find definitive answers to any of these questions; but the questions motivate and guide the design of an interactive robot, which will allow us, and other researchers in the future, to investigate these questions in human-robot interaction. Consequently, it is a key aim of FROG to make the interactive system of the robot (i.e., modules handling, human behaviour sensing and interpretation, dialogue management, and personality-coloured interaction) available to the research community. Benchmarking and quantification of performance will be realised by allowing the wider research community to use our data sets to compare their system's performance with ours. We will search for projects where similar efforts are made to ensure benchmarking opportunities are identified. The performance of our system will be evaluated and quantified by real-world tests with actual users.

Data. Modelling human behaviour depends on having suitable records of human behaviour to learn from. An important aspect of making progress towards socially competent robots therefore lies in providing suitable databases. We consider this aspect to be closely linked to the creation of the interactive system of the FROG robot. Actually, we are facing a chicken-and-egg situation: we need some data for building this system with which we can collect more data, etc. For this reason, the proposal includes an iterative approach to data collection, starting with observing a small number of people outdoors with a camera mounted on a robot that moves around but not on a rough terrain, before working with the various instantiations of the actual system being built. To make this effort valuable for the wide research community beyond the project, FROG aims to make the collected data available to the research community.

Measurable output. In the case of the FROG project, verifying whether the project results have been achieved is rather simple. Hard facts can be used on all relevant aspects:

- Has the navigation system of the FROG robot been built with the capabilities mentioned above? Is it robust? Is it adapting to various environmental conditions and terrain?
- Has the socially competent interactive system been built with the capabilities mentioned above? Is it analysing non-verbal signals in multiple visual modalities, generating appropriate behaviour? Is it running in real time, is it robust?
- Has data been collected, annotated, and released?
- Have various parts of the interactive system been released as (stand-alone) software tools?
- Has the released (compiled) code useful to the community been released?

The milestones in the project have been set to represent various levels of ‘maturity of the system’. It adopts iterative improvements in a spiral life cycle with integration phases. This project is targeting very complex problems involving many stakeholders as well as tools. In addition, there will be a few novel aspects, which are likely to introduce problems which can be solved but are difficult to anticipate. Consequently, it is very unlikely that a classical waterfall life cycle approach to the project organization will be successful in such settings. Namely, due to the complexity of the problem at hand and the intended systems it is very difficult/impossible to extract adequate requirements and foresee all critical issues in the beginning. Instead, requirements will be extracted gradually, as (i) the users are involved in user studies and thorough evaluation and (ii) the developers/researchers experiment with running prototypes/demonstrators in realistic settings, which simply cannot be fully understood through pure introspection or data analysis a priori.

In other words, we cannot provide a perfect plan for the entire duration of the project. In order to avoid significant discrepancies between the planned outcome and the actual results due to unanticipated problems, we are implementing an iterative approach, based on the spiral life cycle, which provides an opportunity for timely corrections.

The iterative nature of the project is reflected in the deliverables of each WP throughout the phases.

The consortium will consider benchmarking at different levels:

- FROG subsystems (from WP1, 2, 3 and 4) will be analysed/compared using available benchmarking tools. Moreover, new data sets will be made available to the community.
- Then, performance testing and benchmark activities for the robot as a whole (WP5)
- Development of outdoor social robot systems evaluation methods and tools

It should be pointed out that the project will need to develop its own evaluation methods, as there are no clear benchmarks in many of the activities: a systematic approach to risk analysis is essential in an ambitious and complex project.