Testing a fully autonomous robotic salesman in real scenarios

Adrián Romero-Garcés
ISIS Group
University of Málaga
Spain, 29071
Email: argarces@uma.es

Luis Vicente Calderita
Robolab Group
University of Extremadura
Spain, 10003
Email: lvcalderita@unex.es

Jesús Martínez-Gómez
Lab. SIMD
University of Castilla-La Mancha
Spain, 02071
Email: Jesus.Martinez@uclm.es

Juan Pedro Bandera
ISIS Group
University of Málaga
Spain, 29071
Email: jpbandera@uma.es

Rebeca Marfil
ISIS Group
University of Málaga
Spain, 29071
Email: rebeca@uma.es

Luis J. Manso
Robolab Group
University of Extremadura
Spain, 10003
Email: lmanso@unex.es

Antonio Bandera
ISIS Group
University of Málaga
Spain, 29071
Email: ajbandera@uma.es

Pablo Bustos
Robolab Group
University of Extremadura
Spain, 10003
Email: pbustos@unex.es

***Abstract*—Over the past decades, the number of robots deployed in museums, trade shows and exhibitions have grown steadily. This new application domain has become a key research topic in the robotics community. Therefore, new robots are designed to interact with people in these domains, using natural and intuitive channels. Visual perception and speech processing have to be considered for these robots, as they should be able to detect people in their environment, recognize their degree of accessibility and engage them in social conversations. They also need to safely navigate around dynamic, uncontrolled environments. They must be equipped with planning and learning components, that allow them to adapt to different scenarios. Finally, they must attract the attention of the people, be kind and safe to interact with. In this paper, we describe our experience with *Gualzru*, a salesman robot endowed with the cognitive architecture RoboCog. This architecture synchronizes all previous processes in a social robot, using a common inner representation as the core of the system. The robot has been tested in crowded, public daily life environments, where it interacted with people that had never seen it before nor had a clue about its functionality. Experimental results presented in this paper demonstrate the capabilities of the robot and its limitations in these real scenarios, and define future improvement actions.**

I. Introduction

The goal of achieving fully autonomous mobile robots, able to operate in populated environments, has been of increasing popularity during the last decade. Over these years, technical developments in this framework have provided robust and reliable systems that have been tested and validated in real-world conditions. Typical scenarios include museums, trade shows and exhibitions [1]. These contexts emerge as a specific application domain of autonomous robots, but also set the robot as a new media technology for exhibition markers and curators.

As K. Arras and W. Burgard pointed out [1], there are very interesting challenges on this new scenario, mainly emanating from the fact that the robot must be able to interact with a non-expert user within a very dynamic environment, where people are speaking and walking around the robot. Within this scenario, the robot must typically be able to speak with one specific person from a group of people, while these people speak and interact among them. Certainly, navigation and perception are challenging tasks. However, the main issue is usually the ability to change the course of action according to the current situation, at human-to-human interaction rates.

In most cases, to afford these challenges, robots are restricted to a specific task (tour-giving, entertainment and animation, education, tele-presence...). They are also limited in their interaction abilities and perceptual capabilities. Despite all these constraints, a fully autonomous robot able to interact with people in daily life environments has to face many issues. It requires many different software components dealing with different tasks, running in parallel and exchanging data. These components are usually organized into a cognitive architecture, that allows them to be simultaneously executed in an efficient and robust way. The cognitive architecture provides all necessary capabilities for performing collaborative tasks: deep representations, domain knowledge and perception and action behaviours [2]. It synchronizes the execution of all the components of the robot to accomplish each use case and learn from experience.

Traditional cognitive architectures separate high-level symbolic planning from geometric plan execution. However, symbolic plans may be slow reacting to changes in the environment. They may also be limited in their efficiency, as they only use data provided by higher-level abstraction layers.

Motivated by human decision-making, the cognitive architecture RoboCog employed in this paper [3] follows the guideline pointed out by Hayes-Roth and Hayes-Roth [4], which showed that humans consider different levels of abstraction in parallel and mentally simulate the execution of the task. Therefore, within RoboCog, action execution, simulation, and perception are intimately tied together, sharing a common motor representation. This inner representation of the outer world is the central module of the architecture for action control. It provides different synchronised interfaces at levels of abstraction that range from the fine-grained aspects to symbolic high level. This central representation helps the robot to be aware of itself, but also to monitor its own capabilities and limitations. The elements of RoboCog are connected to this central representation, and use it to share data at different abstraction levels, to get information about the inner and outer state and to plan next actions.

RoboCog follows a vertical structure, reminiscent of Gat́s three tier architecture [5], with a Hardware Abstraction Layer that provides interfaces to physical subsystems, a layer of behaviors providing perceptual and control skills, and a deliberative planner and executive, in charge of high-level mission unfolding. Task-oriented modules, called compoNets (as they are composed by a set of software components) are connected to the outer world through the Hardware Abstraction Layer. These compoNets process data and connect to the inner representation through specific components called agents. Decision making components that implement the PELEA planner [6] are also connected to this representation and, along with an Executive component based on an Active Grammar-based Model (AGM) [7], allow the robot to perform actions and change its state in the use case (Fig. 3).

The particular implementation of the RoboCog architecture employed in this paper is depicted in Fig. 1. Green upper blocks are low-level action components, pink blocks are low-level perception components, and blue blocks are compoNets in charge of different tasks. It can be seen that behaviors and the deliberative layer communicate through two shared data objects, a kinematic tree (Executive-Geometric in Fig. 1) representing the geometric short-term state of the robot and the environment, and an AGM graph maintaining a symbolic representation of the robot, its environment and the current plan. Both data objects are complementary and together represent the robot belief about itself and the world. The architecture is similar to the one presented in Martínez-Gómez et al. [8], where it is deeply explained.

The rest of the paper is organised as follows: Section II describes the robotic platform employed in these experiments. Section III explains the use case this robot executes within the ADAPTA project. Section IV can be considered the core of the paper. It details the different issues and updates that were confronted during the design and prior tests of the robot, and provides a deep analysis of the final experiments performed in real daily life scenarios. Section V concludes the paper. It discusses its results and describes the future research actions to be performed.

II. Gualzru the robot

The proposed RoboCog architecture has been integrated in the social robot named *Gualzru*. This robot has been implemented in the context of the ADAPTA project (see Section III), and its external appearance has been carefully considered to fit this particular scenario. *Gualzru* aims to be a kind, friendly salesman, that focuses on engaging people in short conversations, and convincing them to follow it to an advertisement panel. It does not require arms as no body gestures are employed in these interactions. It also does not require legs, as it is designed to work in flat floors.

While the external appearance of *Gualzru* is related to the ADAPTA scenario, its internal structure, both hardware and software, has been designed to be as generic as possible. Thus, the objective for this robot is to easily adapt to new applications, scenarios or tasks.

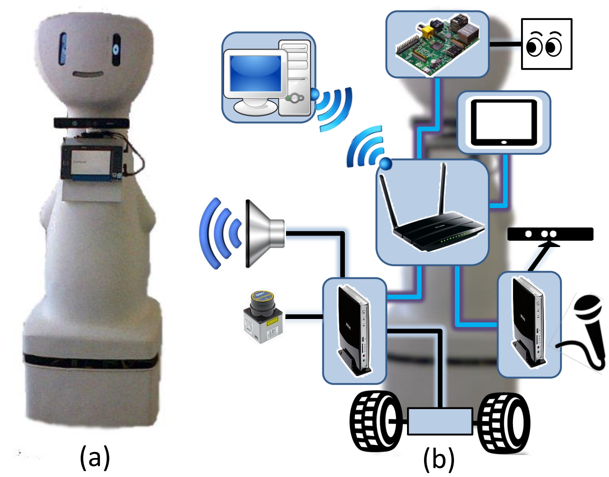

Fig. 2. (a) Gualzru, the robotic salesman; and (b) Hardware structure

Fig. 2 depicts the prototype version of the *Gualzru* robot that has been employed for the experiments presented in this paper. Fig. 2.a shows the external appearance of the robot. It can be seen that the design is simple, and safe, avoiding sharp elements or articulated parts, and hiding the wheels behind the fiberglass case.

Fig. 2.b presents the internal hardware structure of the robot. It is basically composed by a router that connects different processing elements. The main of these elements are a pair of embedded computers. One if this computers runs the WinKinectComp component [8], in charge of processing visual and audio data. It uses the Windows® operating system. This computer is connected to the Kinect® sensor, and to the shotgun microphone employed to capture the voice of the potential client. Fig. 2.a does not include this microphone, that was added to the robot after the preliminary tests described in Section IV (the microphone can be seen in the images shown in Fig. 4).

The other embedded computer uses the Linux operating system and runs the core of the RoboCog software architecture. It is connected to the laser range finder employed for navigation, to the motor controllers that move the robot, and to the speakers.

There are two more processing elements included in the robot. One of them is a Raspberry Pi® device, that is in charge of controlling the eye motion of *Gualzru* and monitoring its battery level. The other is a tablet computer that, for these tests, is only providing information about the state of the robot, or the action it is currently performing (see Fig. 3).

An additional external computer is employed in the system, as Fig. 2.b depicts. This computer is in charge of storing the world representation. It also may run interface modules, that show this inner representation (both geometric and symbolic) and the current state of the robot on screen, and allow the user to start and stop the execution of the use case, or taking direct control of the wheeled robot if required (e.g. dangerous

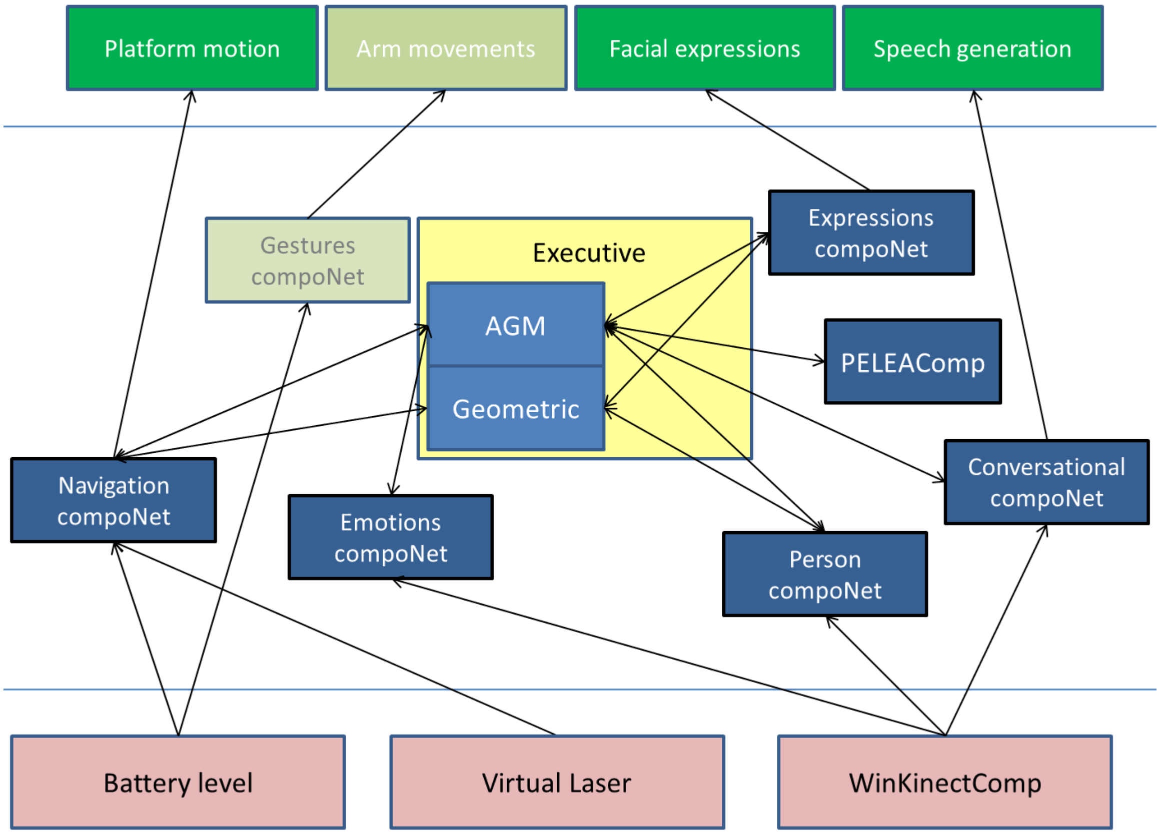

Fig. 1. Implementation of the RoboCog architecture employed in this paper

situation or incorrect behaviour).

III. The ADAPTA scenario

The ADAPTA project ITC-20111030 aims at developing and integrating different technological systems. All these systems focus on providing advertising contents to potential clients in a dynamic, personalized and non intrusive way. Thus, the preferences and needs of each person would be considered for these systems before offering her new products and ads. Some of the different technologies incorporated to achieve this goal are interactive digital signage, holographic representations, and robotics.

In the ADAPTA project, *Gualzru* robot assumes the role of a salesman, that moves in a large shopping area and tries to convince potential clients to follow it to the interactive advertising panel. Fig. 3 shows the use case for this scenario. As depicted, *Gualzru* is firstly waiting on the Starting area (green state in Fig. 3) in the middle of an uncluttered corridor in the Shopping Center. The advertising panel can offer products to any user, thus the robot targets any person in its surroundings without caring about her age range or gender. Once the vision system of *Gualzru* detects a person, it moves towards her. This displacement is very short (3-4 meters maximum) and finishes when the robot approaches the person at social distance (1.5 meters). The robot is continuously monitoring the state of the person while performing this approaching motion. More precisely, it analyzes whether the person faces it or not. If a frontal face is not detected for a certain time, *Gualzru* considers the person is not interested in interacting with it (i.e. she does not look at it) and returns to the Starting area.

If the robot successfully reaches social distance, it introduces itself while classifying the person into a group (using gender and age parameters). Then, it will choose a Product Topic to offer, as Fig. 3 depicts. In the experiments performed in this paper, although the classification process was implemented, its results were not used. Thus, the employed Product Topic was generic, and *Gualzru* focuses just in convincing the person to follow it. The robot is also able to answer certain questions about four topics related to the task: location of the Panel, requested service time, price of the service and extended information requirement. In addition to these topics, it also detects when the user accepts or rejects its invitation. Therefore, comprehension is managed in this work as a classification problem with 4 questions and 2 user decisions (yes-no).

The robot can withdraw from the interaction, once started, for two causes: (i) it is not able to detect the face of the person for a certain time (i.e. the person does not look at it); and (ii) the person rejects its invitation to approach the advertising panel. In any of these cases, *Gualzru* says goodbye and returns to the Starting area. On the other hand, if the person agrees on going to the Panel area, *Gualzru* moves to this area. There, it says the person goodbye and returns to the Starting area. Once it reaches this area, it starts the process again to capture a new target.

The software architecture of *Gualzru* includes a component that continuously checks the batteries. If the batteries level is

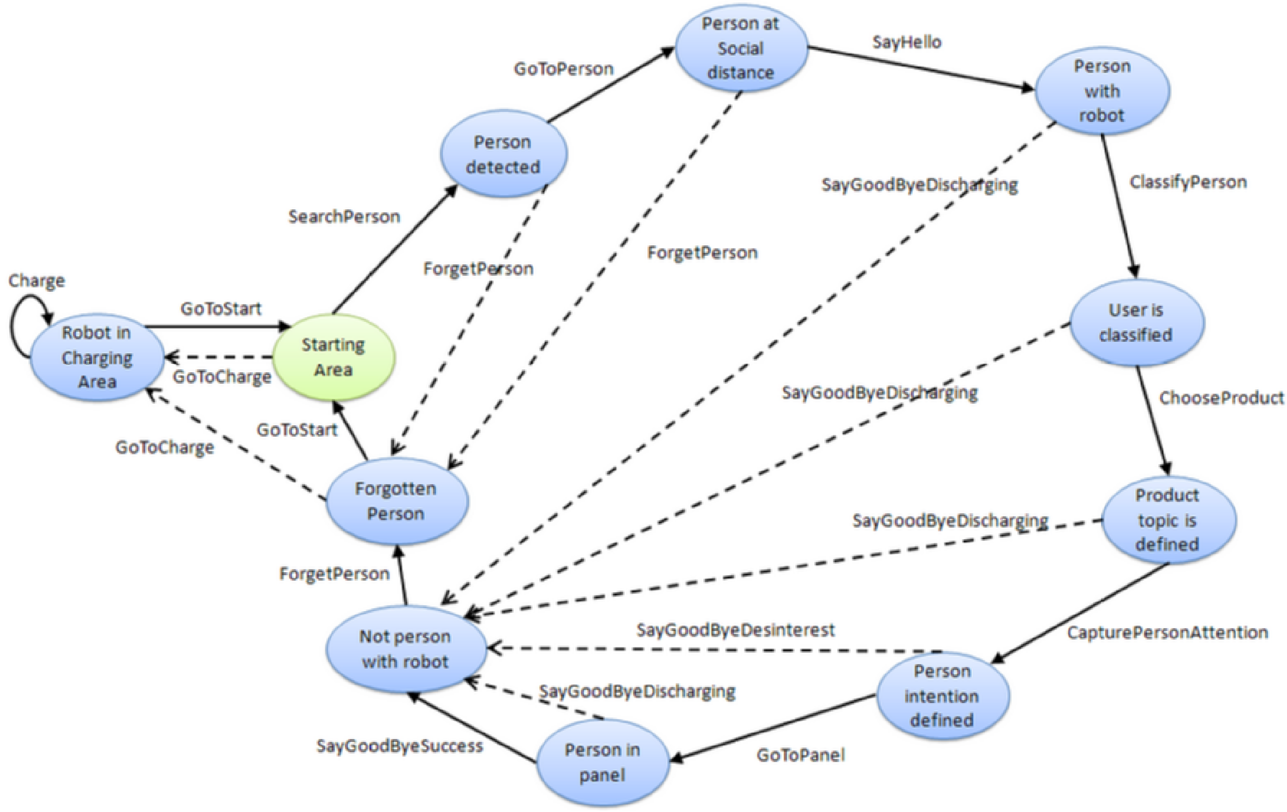

Fig. 3. Use case for the ADAPTA scenario

low, the robot stops the execution of the use case and immediately moves to the Charging area to refill them (GoToCharge transition in Fig. 3). If the robot is maintaining a conversation when this event occurs, it apologizes for leaving before going to the Charging area.

IV. Experiments

Gualzru robot has been tested in different events and group meetings celebrated throughout the last year, in the context of the ADAPTA project. First preliminary tests conducted using a different robot were performed in April, 2014, and they are detailed in Martínez-Gómez et al. [8]. The experimental setup for these tests was a controlled environment in which the robot moved in a restricted area of only 3 square meters.

Complete tests, using the current robotic platform, were carried out in uncontrolled lab environments (a room of about 15 square meters) in June, 2014. By mid-October, 2014 more experiments were performed in which untrained users interacted with the robot in a room. In all cases, *Gualzru* worked in uncontrolled, dynamic environments, populated with people speaking and moving around while taking care of their own business. Most users in these first tests were familiar with the robot as they were involved in the ADAPTA project. Some untrained users were also asked to test the system in order to obtain more valuable feedback, but no extensive tests with unrelated users in shopping environments were conducted.

A. Issues detected in preliminary tests, and performed solutions

Preliminary tests allowed to detect and correct many code bugs, connection problems and technical issues in general. They were also useful to highlight the main capabilities and limitations of the robot [8]. The main of these issues and the addressed solutions are described below.

(i) Following people forever. In the first experiments the robot did not check whether the targeted person was looking at it or not. This produced odd situations, in which a person who passed by the robot walking away was followed by the robot indefinitely, even if the person was clearly not interested in the robot. The current system includes a condition that is checked during the GoToPerson action to avoid this issue: if the robot is not able to detect the face of the person for a certain time period (a good value is 5 seconds) the robot cancels the GoToPerson action and performs instead the ForgetPerson action, returning to the Starting area. The robot will not attend any person while performing this motion. Once it reaches the Starting area, it performs the DetectPerson action again.

(ii) Getting lost after a while. Gualzru uses a local navigator to move around. From the very beginning it was clear that this navigation system, that relies only in odometry to locate the robot, would require additional elements to avoid uncontrolled growing of position errors. Thus, an AprilTag [9] mark was placed close to the panel, and two more marks were located near the Starting area. The positions of these marks are set *a priori* in the world model of the robot. Besides, a component

is added to the software architecture that detects these marks and computes the XYZ position of the robot respect to them. Therefore, the robot can use these marks to relocate itself when they are perceived by the Kinect® RGB camera.

(iii) Unable to correctly hear people in noisy environments. The first implementation of the robot used the array of microphones of the Kinect®. While they worked fairly well in lab environments, they were often unable to understand people in crowded environments, such as a shopping center. Even if the algorithms provided in the Kinect SDK to cancel echo and suppress noise were used, the understanding capabilities of the robot in these environments were very poor. In order to avoid the influence of external audio sources and perceive only frontal audio, a sound diffusion case made of Copopren was added to the prototype. The results were better, but still limited. Finally, the array of microphones of the Kinect device was replaced by an Audio-Technica AT875 Short Condenser shotgun microphone, connected to the computer using an Icicle XLR to USB Mic Converter/Preamp. This device has a narrow radiation diagram that allows it to perceive only audio sources located in front of it. The use of this microphone greatly improved the speech understanding ability of the robot. However, as Fig. 6 depicts, the speech recognition issue is still not solved. The system is still too sensitive to the environmental noise of crowded environments.

(iv) 'No' means 'no'. The conversational system is trained to detect categories using a Bag of Words (BoW) procedure [10] in conjunction with a Bayesian classifier. There is, however, an important issue related to this approach: most users answer the questions of the robot using only a single word, 'yes' or 'no'. But a single word is usually not enough to decide a category using BoW, as that word may be present in many phrases. This issue is specially relevant for the 'no' answer, as conditional 'no' is frequently employed. Thus, it was necessary to create a shortcut in the grammar for these monosyllabic responses. In the final version of the conversational module, these words directly trigger the respective response without being processed by the BoW procedure.

(v) Facing the person during the conversation. In order to offer more natural interactions and better track the person, a behaviour was incorporated to the navigation component, that simply makes the robot turn to face the human during the conversation.

(vi) Near mode required. When *Gualzru* works in crowded, noisy scenarios, people sometimes do not hear it properly (see Fig. 6). These people will approach the robot to understand what it is saying. A standard Kinect® device will stop tracking the human if she is closer than about one meter. The Kinect for Windows® device, however, incorporates a 'Near' mode that allows tracking upper body motion at close range. The WinKinectComp component [8] was modified to switch between 'Default' and 'Near' modes to prevent this issue. These modifications improve the results. The question, however, remains open, as some people move so close to the robot that even using this feature they are not perceived by the sensor.

(vii) Slow approach, fast retreat. The motion speed of *Gualzru* was tuned, thus it moves slower when approaching people, and faster when going back to the Starting, Panel or Charging areas.

Fig. 4. Images taken during the experiments performed at the University of Málaga

B. Experimental setup

On December, 2014, the current version of *Gualzru* robot was tested in a real working scenario. The system was deployed in the hall of the Escuela de Ingenierías, at the University of Málaga. The area where the robot was employed was about 70 square meters. Fixed obstacles included a column and some tables, but most of the area was free for the robot to move. The hall was populated by students and the tests extended for two mornings. The robot operated without human intervention, and engaged people who passed near it. These people had no *a priori* knowledge about the robot nor its functionality. Fig. 4 shows the experimental setup and different images taken during the experiments.

Fifty people who interacted with the robot were asked to fill a questionnaire after the experience. The questionnaire is designed as a Likert scale, although it uses six levels, from 0 to 5, to remove the neutral option (middle point). It is similar to that employed by Joosse et al. [11] to generate the database BEHAVE-II. Its main difference is that it has been created not from the point of view of the person observing the behavior of the user against the presence of the robot, but from the point of view of the same user that interacts with the robot. In this sense, we can consider that collects influences of questionnaires of the Almere original model or the man-machine interaction. In particular, the questionnaire includes a collection of questions arranged in four blocks (navigation, conversation, interaction and general sensations). The user fills the questionnaire giving a value for each response between 5 (completely agree) and 0 (completely disagree). These questions are listed in Table I.

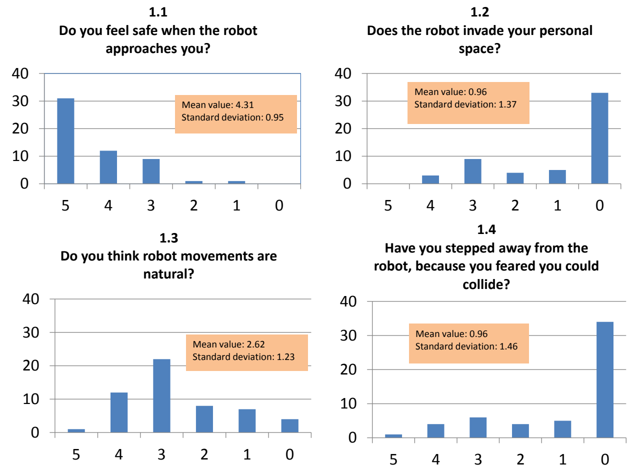

Fig. 5. Questionnaire results: Navigation

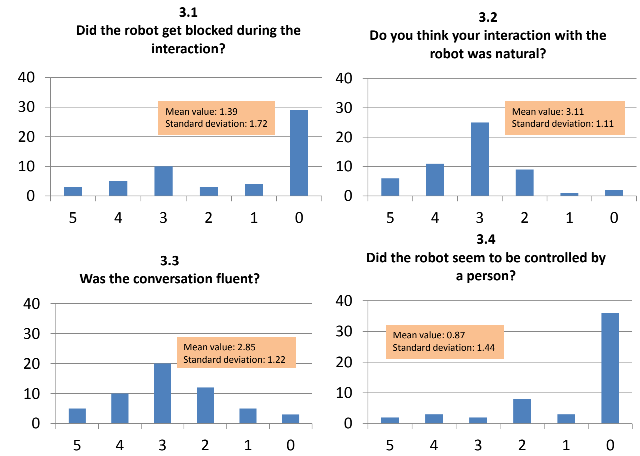

Fig. 7. Questionnaire results: Interaction

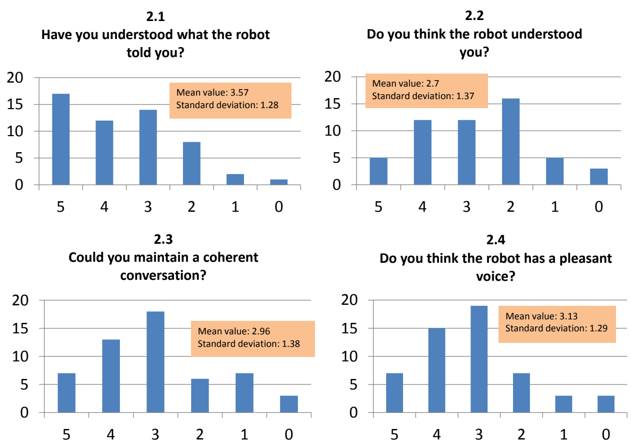

Fig. 6. Questionnaire results: Conversation

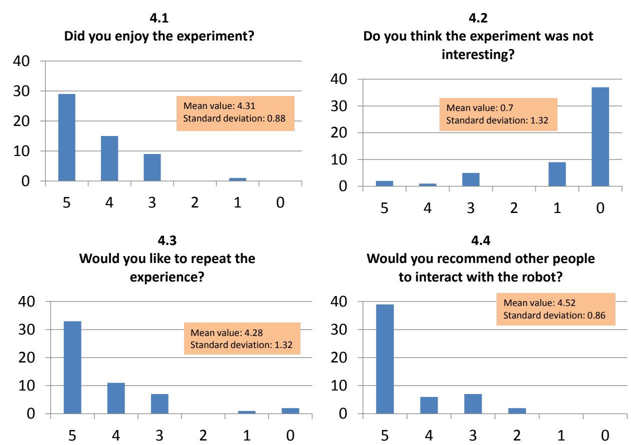

Fig. 8. Questionnaire results: Overall sensations

TABLE I. QUESTIONNAIRE RESULTS (50 TESTS)

Question	$\bar{x}$	σ
1.1 Do you feel safe when the robot approaches you?	4.31	0.95
1.2 Does the robot invade your personal space?	0.96	1.37
1.3 Do you think robot movements are natural?	2.62	1.23
1.4 Have you stepped away from the robot?	0.96	1.46
2.1 Have you understood the robot?	3.57	1.28
2.2 Has the robot understood you?	2.70	1.30
2.3 Was the conversation coherent?	2.96	1.38
2.4 Do you like the voice of the robot?	3.13	1.29
3.1 Did the robot get blocked?	1.39	1.72
3.2 Was the interaction natural?	3.11	1.11
3.3 Was the conversation fluent?	2.85	1.22
3.4 Did the robot seem to be tele-operated?	0.87	1.44
4.1 Did you enjoy the experiment?	4.31	0.88
4.2 Do you think the exp. was not interesting?	0.70	1.32
4.3 Would you like to repeat?	4.28	1.32
4.4 Would you recommend it to other people?	4.52	0.86

C. Questionnaire results

Figs. 5, 6, 7 and 8 show the values obtained in the questionnaires. Table I lists the mean values and standard deviations for each of the questions.

Navigation results, depicted in Fig. 5 show that the robot is perceived as a safe device, although its movements are not really natural. This issue can be considered a minor drawback, as it does not reduce the efficiency of the robot nor its ability to capture the attention of the people. In any case, it would be interesting to improve the algorithms employed to set the speed of the robot. New algorithms will perform third order spline interpolation to change speed values, instead of the currently employed linear interpolation. This change should produce smoother and more natural speed changes [12].

Fig. 6 shows that the conversational system is the weak point of the robot. Some people did not correctly understood the robot due to the environmental noise, and the voice of the robot was perceived as not particularly pleasant. But the most important issue was related to the understanding capabilities of *Gualzru*. Even when using the shotgun microphone these capabilities were strongly limited. The system is too sensitive to environmental noise and echos, and it gets also confused when there are several people speaking around the robot (see Fig. 4). This situation is more common than expected due to the interest the robot produces. Additional issues such as different accents, voice volumes, etc add more difficulties to the scenario.

The interaction results (Fig. 7) show that the robot does not get blocked, and it is perceived as fully autonomous. But the limited conversational abilities of *Gualzru* influences its interaction capabilities.

Despite these limited conversational skills, *Gualzru* achieves its objectives, as Fig. 8 shows. It catches the attention of people. Most of them enjoyed the experiment, would recommend the experience and would like to repeat it.

The comparison of these results against the ones collected in the first experiments (detailed in Martínez-Gómez et al. [8]) reveals that successive updates in the robot have made it more robust, and its conversational abilities, while still constrained, have been significantly improved.

V. Conclusion

The experiments presented in this paper move a social robot from lab environments, or controlled tests, to a real daily life scenario. People populating this scenario (that can be really crowded sometimes, as Fig. 4 depicts) have no previous idea about the robot functionality and abilities. Experiments lasted for three mornings, and more than 50 people interacted with the robot. Filled questionnaires show that they liked the experience, and that the robot was perceived as safe and interesting.

Gualzru worked autonomously during all these experiments. No human intervention was performed. It was able to complete the use case and reacted coherently to nearly all situations. Nevertheless, some minor improvements are going to be addressed in the navigation system to produce smoother and more natural motion. On the other hand, while AprilTag marks are adequate to relocate the robot, they require direct visual contact with the robot. Some partners in the ADAPTA project are working in RFID location, and future work will address the inclusion of such elements to help *Gualzru* locate the Starting and Panel area.

The main issue for *Gualzru* is related to its limited conversational abilities. These limitations represent a severe drawback for its performance. After several changes that involved different microphones and algorithms, 50% of the people that interacted with the robot in these real scenarios think that it is able to maintain a coherent conversation. We think that this is not enough for a robust, useful robot. Our future work will mainly focus in this topic. But the speech recognition issue may be hard to solve in noisy, crowded environments in which even people find difficulties in understanding each other. Thus, our idea is to look for alternative methods to allow people communicate with the robot. More precisely, speech recognition will be reinforced with a more active use of the tactile screen installed in the chest of the robot. Robot phrases will be displayed in this screen, and it will be possible for the person to answer the robot by touching it. Therefore, *Gualzru* will retain its conversational abilities but new interfaces will be offered to increase its robustness and usefulness.

Acknowledgment

This paper has been partially supported by the Spanish Ministerio de Economía y Competitividad TIN2012-TIN2012-38079-C03 and FEDER funds, and by the Innterconecta Programme 2011 project ITC-20111030 ADAPTA.

References

[1] K. Arras and W. Burgard, *Robots in Exhibitions*. Lausanne, Switzerland: EPFL, 2002.

[2] M. Beetz, D. Jain, L. Mosenlechner, and M. Tenorth, "Towards performing everyday manipulation activities," *Robotics and Autonomous Systems*, vol. 58, no. 9, pp. 1085–1095, 2010.

[3] P. Bustos, J. Martínez, I. García, L. Rodríguez, P. Bachiller, L. Calderita, L. J. Manso, A. Sánchez, A. Bandera, and J. Bandera, "Multimodal interaction with loki," in *XIV Workshop of Physical Agents (WAF 2013)*, Madrid, Spain, September 2013, pp. 53–60.

[4] B. Hayes-Roth and F. Hayes-Roth, "A cognitive model of planning," *Cognitive Science*, vol. 3, no. 4, pp. 275–310, 1979.

[5] E. Gat, "On three-layer architectures," *Artificial intelligence and mobile robots*, pp. 195–210, 1998.

[6] E. Quintero, V. Alcázar, D. Borrajo, J. Fernández-Olivares, F. Fernández, A. G. Olaya, C. Guzman, E. Onaindia, and D. Prior, "Autonomous mobile robot control and learning with the pelea architecture," in *Proc. Automated Action Planning for Autonomous Mobile Robots (PARM 2011)*, San Francisco, EEUU, August 2011, pp. 51–56.

[7] L. Manso, "Perception as stochastic sampling on dynamic graph spaces," Ph.D. dissertation, Univ. of Extremadura, Spain, 2013.

[8] J. Martínez-Gómez, R. Marfil, L. V. Calderita, L. J. Manso, A. Bandera, A. Romero-Garcés, and P. Bustos, "Toward social cognition in robotics: Extracting and internalizing meaning from perception," in *XV Workshop of Physical Agents (WAF 2014)*, León, Spain, June 2014, pp. 93–104.

[9] E. Olson, "Apriltag: A robust and flexible multi-purpose fiducial system," University of Michigan APRIL Laboratory, Tech. Rep., May 2010.

[10] H. M. Wallach, "Topic modeling: beyond bag-of-words," in *Proc. of the 23rd international conference on Machine learning*, 2006, pp. 977–984.

[11] M. Joosse, A. Sardar, M. Lohse, and V. Evers, "Behave-ii: The revised set of measures to assess users' attitudinal and behavioral responses to a social robot," *International Journal of Social Robotics*, vol. 5, no. 3, pp. 379–388, 2013.

[12] Q. Huang, K. Yokoi, S. Kajita, K. Kaneko, H. Arai, N. Koyachi, and K. Tanie, "Planning walking patterns for a biped robot," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 280–289, 2001.