Personalized short-term multi-modal interaction for social robots assisting users in shopping malls

Luca Iocchi[1], Maria Teresa Lázaro[1],
Laurent Jeanpierre[2], Abdel-Illah Mouaddib[2]

[1] Dept. of Computer, Control and Management Engineering
Sapienza University of Rome, Italy.
[2] GREYC, University of Caen Lower-Normandy, France

Abstract. Social robots will be soon deployed in large public spaces populated by many people. This scenario differs from personal domestic robots, since it is characterized by multiple short-term interactions with unknown people rather than by a long-term interaction with the known user. In particular, short-term interactions with people in a public area must be effective, personalized and socially acceptable. In this paper, we present the design and implementation of an Human-Robot Interaction module that allows to personalize short-term multi-modal interactions. This module is based on explicit representation of social norms and thus provides a high degree of variability in the personalization of the interactions, maintaining easy extendibility and scalability. The module is designed within the framework of the `COACHES` project and some implementation details are provided in order to demonstrate its feasibility and capabilities.

1 Introduction

The new challenge for robotics in the near future is to deploy robots in public areas (malls, touristic sites, parks, etc.) to offer services and to provide customers, visitors, elderly or disabled people, children, etc. with increased welcoming and easy to use environments.

Such application domains present new scientific challenges: robots should assess the situation, estimate the needs of people, socially interact in a dynamic way and in a short time with many people, exhibit safe navigation and respect the social norms. These capabilities require the integration of many skills and technologies. Among all these capabilities, in this paper we focus on a particular form of Human-Robot Interaction (HRI): Personalized Short-term Multi-Modal Interactions. In this context, *Personalized* means that the robot should use different forms of interactions to communicate the same concept to different users, in order to increase its social acceptability; *Short-term* means that the interactions are short and focused on only one particular communicative objective, avoiding long and complex interactions; while *Multi-modality* is obtained by using different interaction devices on the robot (although in this study, we focus only on speech and graphical interfaces).

The solution described in this paper is developed within the context of the `COACHES` project[1], that aims at developing and deploying autonomous robots

[1] Funded by MIUR Italy and ANR France within the CHIST-ERA research program.

providing personalized and socially acceptable assistance to customers and shop managers of a shopping mall. The main contribution of this paper is on the architecture of the Human-Robot Interaction module that has several novelties and advantages: 1) integrated management of all the robotic activities (including basic robotic functionalities and interactions) through the use of Petri Net Plans, 2) explicit representation of social norms that are domain and task independent, 3) personalized interactions obtained through explicit representation of information and not hand-coded in the implementation of the robot behavior.

In the rest of this paper, after an analysis of the literature in personalized human-robot interaction (Section 2) and a brief description of the general architecture of the `COACHES` system (Section 3), we describe the human-robot interaction component and, in particular, our approach to personalized short-term multi-modal interaction (Section 4). In Section 5, we provide some examples of application of the proposed system and finally we draw conclusions in Section 6.

2 Related Work

The use of service robots interacting daily with people in public spaces or workplaces has become of increased interest in the last years. In this context, the development of the robotic system should focus on creating confortable interactions with the people the robot has to share its space.

Gockley *et al.* [3] showed that people usually express more interest and spend more time during the first contact with the robot. However, after the novelty effect, the time of the interactions decreases which suggests people's preference for *short-term* interactions.

In order to address this decrease in the people engagement, Lee *et al.* [4] demonstrated in a 4-month experiment that *personalized* interactions allow to maintain the interest of the users over time. The experiment consisted of a robot delivering snacks in a workplace and the personalization was carried out through customized dialogues where the robot addressed the users by their names and commented the users' behaviours like their frequency of usage of the service or their snack choice patterns. Conversely, in certain contexts like in rehabilitation robotics, it is desired to have longer interactions with the patient, so the robot can assist and encourage him to do his exercises. In [8], it is shown how adapting the robot behaviour to the patient personality (introvert or extrovert) increases the level of engagement in the interaction.

Certain works aim at personalizing the interaction by learning from its user. For example, in [5] a certain task is commanded to the robot which receives a feedback from the user if the final state of an action is desirable for him. In this way, the robot learns from the user's preferences which are registered in a user profile. With this knowledge and the feedback it keeps receiving from the user, the robot can anticipate his needs and pro-actively act to fullfil his needs. In [6] the robot adapts its behaviour defined by parameters like the distance to the person or its motion speed, among others, using a reinforcement learning technique where feedback from the user is given subsconciously through body signals read directly from the robot sensors.

In contrast to these works, our approach for personalized human-robot interaction is not based on learning the personality of the user, but on a set of *social*

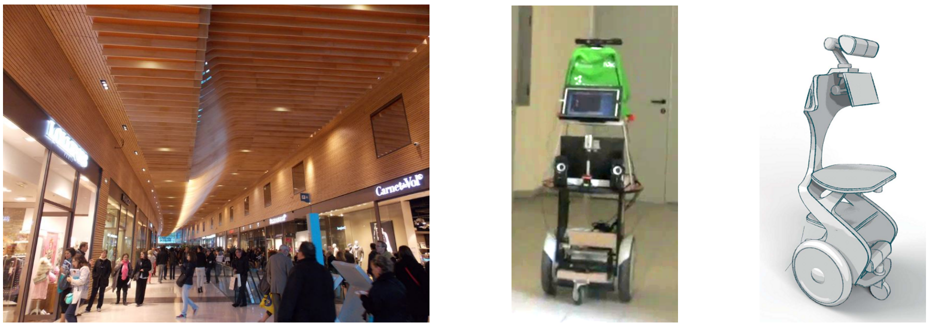

Fig. 1: `COACHES` environment and robots.

norms that are present in our everyday lifes in human interactions. Moreover, our architecture is designed on domain and task-independent representation of information, providing for a high variability of personalized behaviors, with a simple declarative definition of the social norms that we want the robot to apply. This provides many advantages in terms of extendibility and scalability of the system. The proposed approach extends a previous work [7] about the design of social plans, by adding the notions of user profiles and of personalized interactions.

3 COACHES Environment, Hardware and Software Architecture

In the `COACHES` project (October 2014 - September 2017), we aim to study, develop and validate integration of Artificial Intelligence and Robotics technology in order to develop robots that can suitable interact with users in a complex large public environment, like a shopping mall. Figure 1 shows, on the left, the *Rive de l'orne* shopping mall in Caen (France) where the experimental activities of the project will be carried out, on the middle a prototype of the robot that will be used for the preliminary experiments and, on the right, the design of the robots that will be realized in Fall 2015.

As shown in the figure, in contrast with previous work on social robotics and human-robot interaction, the `COACHES` environment is very challenging, because it is populated by many people and the robot is expected to interact with many unknown and non-expert users. Moreover, we aim at a more sophisticated interaction using multiple modalities (speech, gesture, touch user interfaces) and dialog generated on-line according to the current situation and the robot's goals. Although these characteristics are not completely new in related projects, we believe that the `COACHES` project will provide important insights for actual deployment of intelligent social robots in large populated public areas.

The software architecture of the `COACHES` robots is shown in Figure 2. The architecture comprises a typical configuration of an autonomous robot where all the decisions are made on-board based on sensors available.

An open architecture (hard/soft) and standard technologies available will be used, so that it will be easy to extend and/or adapt the capabilities of the

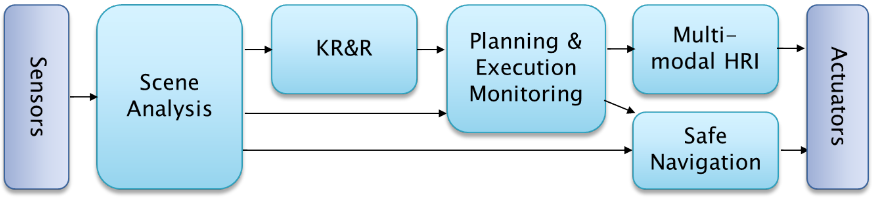

Fig. 2: `COACHES` software architecture

system during the whole length of the project (especially to integrate and test various algorithms and/or sensors). Such an open architecture will also simplify and optimize integration efficiency as well as re-use of assets in other projects or products.

For the development of the robotic software components, the Robot Operating System (ROS)[2], which is the standard middleware for robotics applications, has been selected. ROS provides the middleware infrastructure to effectively share information among the many modules implementing various functionalities on each robot. Moreover, our software architecture includes an interface[3] for sharing information among the robots and between ROS and non-ROS components of the system, based on serializing and deserializing ROS messages as strings sent over TCP.

The main software components that are under development for control, reasoning and interaction functionalities of the system are briefly described below.

- *Scene analysis* includes sensor processing procedures for both on-board robot devices and static sensors in the environment in order to determine the current situation and understand events that are of interest for the system.
- *Knowledge-based representation and reasoning* defines the formalism and the procedure to represent and reason about the environment and the task of the robots. It provides the goals that the robots should achieve given the current situation.
- *Planning and execution monitoring* generates the plans to achieve the desired goals and monitor their execution for robust behaviors.
- *Multi-modal HRI* defines a set of modalities for human-robot interaction, including speech recognition and synthesis, touch interaction, graphical interface on a screen mounted on the robot and Web interfaces.
- *Safe navigation* guarantees safety movements and operations of the robot in a populated environment.

In the next section, we focus on the description of the *Short-Term Multi-Modal HRI* module and, in particular, we show our approach to personalized interactions with users of the shopping mall. Although at this time all the other modules have not been fully realized, a minimum set of functionalities needed to test the HRI component are present.

[2] www.ros.org
[3] https://github.com/gennari/tcp_interface

4 Personalized Short-term Multi-Modal Interactions

As already mentioned, one of the main goals of the `COACHES` project is personalized short-term multi-modal interactions with non-expert users, that are typical customers of a shopping mall.

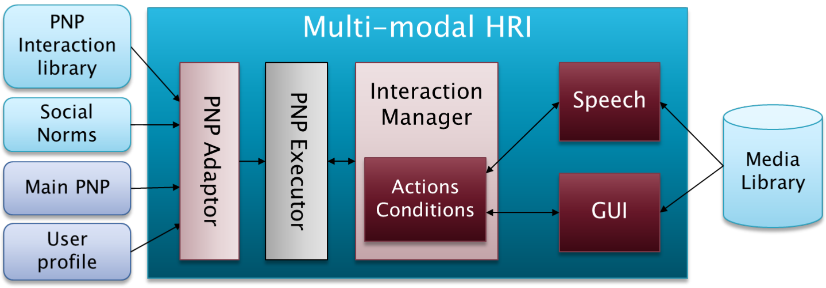

Fig. 3: Architecture of Human-Robot Interaction module.

The architecture of the HRI module is illustrated in Figure 3. Data available to this module are Petri Net Plans (PNP) encoding the desired behavior of the robot, social norms, a user profile and a multi-media library.

The PNPs (as described later) encode the overall behavior of the robot, as generated by the planning and reasoning components of the system. The behavior include both basic robotic actions (e.g., moving in the environment) and interaction action. The user profile is the information available about the user that is interacting with the robot. Among acquisition means for user profiles, it is possible to think about users wearing an RFID tag containing personal information read by an RFID reader on-board the robot, or to the request of swiping a fidelity card, enter a personal password or showing to the robot a QR-code, in order to communicate to the robot the user profile. In our implementation, we have used a simple identification mechanism based on recognizing QR-codes shown by the user to the robot on-board camera.

Finally, the media library is a collection of multi-media data (text, images, animations, video, etc.) that are linked to the communication activities of the robot and to the user profiles. We assume that in this library there are different versions of the same communication target for different users. For example, ice-cream advertisement can have a different spoken text and different displayed images or videos for children and adults.

In the remaining of this section, we will describe in more details the components of this module.

4.1 PNP Adaptor and Executor

The HRI module is implemented within the framework of the Petri Net Plans (PNP) formalism [9]. PNPs are based on two main concepts: *actions* (i.e., output operations) and *conditions* (i.e., input operations). Actions include motion of the robot in the environment, spoken sentences issued by the on-board speakers, text, images, videos or animations shown on the on-board screen, etc. Conditions include the result of perception routines (e.g., image processing or speech recognition), the input given by a user through a GUI on the on-board screen, information about personal data of user acquired through a reader of fidelity cards, etc.

The use of PNPs for representing in an integrated way all these different kinds of actions and conditions allows for a strong coordination between the different sub-systems of the robot and for showing more complex behaviors and, in particular, a multi-modal interaction that can be easily customized according to the user.

The main plan, which includes *interaction plans* for HRI behaviors, generated by the reasoning and planning sub-system, is first processed by the PNP Adaptor and then executed by the PNP Executor. Both these modules are general-purpose, since all the relevant information is provided by external sources with an explicit representation. More specifically, the PNP Adaptor generates a personalized plan, given a main plan, a library of interaction plans, a set of social norms, and the user profile. The generated personalized plan is then executed by the PNP Executor.

PNP Adaptor is implemented through an algorithm that transforms the Main PNP and the associated Interaction PNPs according to the social norms applied to the specific user profile. More specifically, the input plan is composed by a user-generic Main PNP that calls Interaction PNPs as sub-routines. All these plans are processed and transformed by applying the social norms (described as rules) customized to the current user profile.

The social norms are domain and task independent and are represented using a propositional logic formalism that follows the one described in [2]. Given a set of propositions U related to user profiles and a set of propositions I related to forms of interactions, and given the set of formulas U^* over U and the set of literals I^+ over I, a social norm is represented as a pair $(\phi, \delta) \in U^* \times I^+$, with the meaning that if ϕ is true, then δ is mandatory (i.e., it must occur), or, in other words, $\neg\delta$ is forbidden.

Some examples of social norms implemented in our system and considered in the examples in the next section are illustrated in Table 1.

Given a set of social norms S and a user profile u from which it is possible to determine the truth of the formulas in U, then it is possible to derive all the literals in I^+ that are implied by the social norms and u. In other words, it is possible to compute the set of propositions Δ_u such that $S \wedge u \models \Delta_u$. These propositions can be seen as the personalization of S to u. For example, if the user profile u satisfies elder and deaf, Δ_u contains { use_big_font, display_spoken_text, use_simple_GUI, ¬ use_speech }. In this paper, we do not explicitly consider the case in which Δ_u may become inconsistent. Of course, several mechanisms could be implemented for solving this issue, such as adding preferences or priorities to propositions.

The personalized propositions Δ_u affect the execution of the output actions of the HRI module. Each action in the PNPs is personalized by adding the appropriate propositions as arguments. As described later in the section, in this paper we consider two kinds of output interaction actions: *Say*, related to the

(child, use_animation)
(elder, use_big_font)
(elder, use_simple_GUI)
(deaf, ¬ use_speech)
(blind, ¬ use_display)
(elder ∨ deaf, display_spoken_text)
(elder ∨ deaf ∨ blind, ask_for_guidance)
(blind, use_detailed_speech)
(blind, notify_guidance)
(first_time_user, detailed_instructions)
(¬ first_time_user ∧ young, ¬ detailed_instructions)
(child ∨ very_young, ¬ use_baby_care_room)
(foreign, speak_English)

Table 1: Domain-independent social norms.

Speech module, and *Show*, related to the Graphical Interface module. Therefore, literals associated with *Say* (e.g., ¬ use_speech) are added as parameters of all the actions *Say* in the PNPs, while literals associated with *Show* (e.g., use_big_font, display_spoken_text, use_simple_GUI) are added as parameters of all the actions *Show* in the PNPs. These parameters determine the personalized interaction and will be considered by the Interaction Manager.

PNP Executor is a general-purpose executor of PNP already described in [9] and successfully used in many applications. PNP Executor treats actions and conditions without giving them any semantics and controls only the flow of execution. The actual execution of the basic actions and conditions is responsibility of the Interaction Manager.

4.2 Interaction Manager

The interactions are coordinated by an Interaction Manager (IM), which manages all the robot activities (both the ones related with human-robot interaction and the ones used for implementing the basic robotic functionalities). Its goal is thus to provide effective robot behaviors, including the personalized short-term multi-modal interactions described in this paper.

The IM is an action and condition server that executes actions and provides conditions, according to the requests of the PNP Executor module. It thus includes the definition of a set of primitive actions and conditions that are activated according to the plan under execution. For the interaction behavior, actions and conditions are actually related to the Speech and Graphical Interface (GUI) modules described later. While the actions and the conditions related to the basic robot abilities (such as navigation, localization, perception, etc.) are not illustrated and described here, since the focus of this paper is on interaction. The IM is also responsible to activate actions according to the personalized parameters defined by the PNP Adaptor module.

4.3 Speech and Graphical Interfaces

The interaction modalities considered so far in the project are speech and graphical interfaces.

Speech recognition and synthesis. The speech component allows the robot to communicate with humans through vocal interactions. It is formed by Automatic Speech Recognition (ASR) and Text-To-Speech (TTS).

The ASR component analyzes audio data coming from a microphone and extract semantic information about the spoken sentences, according to a predefined grammar. This component allows the robot to understand user spoken commands. The speech recognition module is based on the Microsoft engine and on a further processing module that builds the semantic frames of the recognized sentences. More details on the approach are available in [1].

The TTS component transforms text messages in audio data that are then emitted by the speakers on-board the robot. This enables the robot to speak to people. The Microsoft TTS engine is used for this module.

Graphical User Interface. The GUI component implements a graphical input and output interface between users and robots that is displayed through the touch screen on-board the robot. The GUI defines actions (i.e., output operations) and conditions (i.e., input operations) that are integrated in the IM with other communication primitives (e.g., speech) in order to implement a multi-modal interaction.

The Speech and GUI components make available to the IM the implementation of actions and conditions that are executed according to the PNPs. These are summarized in the following table.

	Action	**Condition**
Speech	*Say* speak information though TTS	*ASR* Results of ASR
GUI	*Show* show information on the GUI	*GUI* Results of GUI input

The actions implemented at this level are parametric with respect to a set of parameters expressed as propositions and used to define the social norms. As mentioned above, during the process, general actions are associated to specific parameters depending on the user profile. This parameters are considered to specialize the execution of the actions. Two kinds of specializations are considered: 1) modification of some internal parameters of the action (for example, the size of the font in a displayed text), 2) selection of the proper media to communicate. The second specialization is related to the presence of multiple options in the Media Library for the same communicative target. In these cases, each option is labeled with a precondition using the same interaction propositions in I. Therefore, it is possible to select appropriate media considering the personalized propositions Δ_u.

5 Examples of personalized interactions

In this section we will show through a set of examples how general purpose social norms are used to affect the behavior of the robot in a declarative way. The examples are taken from the use cases of the COACHES project and they will be eventually fully implemented and tested with real robots in the shopping mall in Caen. The examples below refer to the social norms described in Section 4 and assume user profiles are available.

Example 1. Advertising. One of the tasks of the `COACHES` robot is to show advertisements to users of the shopping mall. These advertisements (in forms of text, images, videos, etc.) are provided by the shop manager and stored in the Media Library. In one form of advertising planned in the project the robot knows the user profile. In this case the Interaction Module described in the previous section can activate personalized messages. Effects of personalized interactions in this example are: i) animation instead of videos for children, ii) big fonts and simple GUI for elderly people, etc.

Example 2. Directions and guiding. The robot is able to give directions and guide people in the mall. Requests are acquired either by voice or graphical interface and the robot uses its semantic map of the environment to show directions or accompany the user. In this case the following personalized behaviors can be obtained: i) for elderly people, a simple GUI shows the direction; ii) the interaction with a deaf and elder person is made with graphical interface only; iii) the interaction with a blind person uses only voice. In all the three cases, the robot offers to accompany them and for the blind person a special notification is given with instructions of how the guidance will happen.

Example 3. Baby care rooms. Baby care rooms can be used by parents, but must be reserved and they are locked when not in use. The robot can enable this service upon request. Some personalized interactions in this case are: i) a new user is fully instructed with detailed instructions about how to use the service; ii) a user that already used the service a few time ago is given directly the access to the baby care room; iii) children or very young users will be notified that they are not allowed to use the service.

Notice that all these examples are implemented without explicit coding the corresponding behaviors. The expected personalized behavior is the effect of the application of the social norms to the user profile and of the corresponding modifications of the plans that activate actions with proper parameters. Notice also that the social norms are not specific of any particular task. This allows for a high level of extendibility. For example, adding, removing or modifying social norms allow for a significant change of behavior of the robot with different users without requiring any change (or just minimal changes) in the implementation of the actions. For example, assuming that we want to add the capability of the robot to regulate the volume of its voice and to personalize this feature. With our architecture it is sufficient to do the following steps: add a parameter about volume in the *Say* action (e.g., corresponding to a new proposition loud_speech in I) and a social norm (elder, loud_speech) in S. All the interactions with elder people now will use an increased volume of the robot speech.

6 Conclusions

In this paper we have presented our architecture for personalized short-term human-robot interaction to be used by `COACHES` robots that will autonomously provide services to customers in a shopping mall. Robot actions and interactions with users have been described through PNPs that can be dynamically adapted according to the user profile and a set of domain-independent social norms. This capability provides the system with a high level of scalability and, as shown in our examples, allows for being easily extended to a variety of interactions. Implementation of the HRI module presented in this paper has been tested in our lab with a prototype robot and not yet in the real environment with real users.

The on-going `COACHES` project is the main experimental test-bed for the work presented in this paper. Future work will thus include a user study, whose main focus will be assessing improved acceptance of a social robot with personalized interactions. With the approach described in this paper, producing different versions of the interaction behavior of the robot is as easy as adding or removing a social norm. However, in certain cases, personalization based on social norms may not be sufficient due to individual exceptions. Therefore, a more detailed individualization level applied to single users will also be subject of further studies.

References

1. Bastianelli, E., Castellucci, G., Croce, D., Basili, R., Nardi, D.: Effective and robust natural language understanding for human-robot interaction. In: Proceedings of 21st European Conference on Artificial Intelligence. pp. 57–62 (2014)
2. Boella, G., Pigozzi, G., van der Torre, L.: Normative framework for normative system change. In: Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS) (2009)
3. Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A.C., et al.: Designing robots for long-term social interaction. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems. pp. 1338–1343 (2005)
4. Lee, M.K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J., Savetsila, S.: Personalization in HRI: A longitudinal field experiment. In: ACM/IEEE Int. Conf. on Human-Robot Interaction. pp. 319–326. New York, NY, USA (2012)
5. Mason, M., Lopes, M.: Robot self-initiative and personalization by learning through repeated interactions. In: ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI). pp. 433–440 (2011)
6. Mitsunaga, N., Smith, C., Kanda, T., Ishiguro, H., Hagita, N.: Adapting robot behavior for human–robot interaction. IEEE Transactions on Robotics 24(4), 911–916 (2008)
7. Nardi, L., Iocchi, L.: Representation and execution of social plans through human-robot collaboration. In: Beetz, M., Johnston, B., Williams, M.A. (eds.) Social Robotics, Lecture Notes in Computer Science, vol. 8755, pp. 266–275. Springer International Publishing (2014)
8. Tapus, A., Ţăpuş, C., Matarić, M.J.: User-robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. Intelligent Service Robotics 1(2), 169–183 (2008)
9. Ziparo, V.A., Iocchi, L., Lima, P.U., Nardi, D., Palamara, P.F.: Petri net plans - A framework for collaboration and coordination in multi-robot systems. Autonomous Agents and Multi-Agent Systems 23(3), 344–383 (2011)