Spoken Communication with CAMBADA@Home Service Robot

Ciro Martins[1,2], António Teixeira[1], Eurico Pedrosa[1] and Nuno Almeida[1]

[1] Department Electronics, Telecommunications & Informatics/IEETA
Aveiro University, Aveiro, Portugal
[2] School of Technology and Management
Aveiro University, Águeda, Portugal

Abstract. Spoken language is a natural way to control the human-robot interaction, especially for mobile service robots. It has some important advantages over other communication approaches: eyes and hands free, communication from a distance, even without being in line of sight and no need for additional learning for humans. In this paper, we present the spoken dialog framework integrated in our mobile service robot CAMBADA@Home, a robotic platform aimed at move into a living space and interact with users of that space. The proposed framework comprises three major spoken and natural language processing components: an Automatic Speech Recognition component to process the human requests, a Text-to-Speech component to generate more natural responses from the robot side, and a dialog manager to control how these two components work together.

1 Introduction

Service robots have the potential to enhance the quality of life for a broad range of users. In fact, developing different types of robotic systems that can be able to interact with the human world [1] is one major challenge for the 21st century in the robotics area. It is expected that Robotics play a key role when targeting social challenges such as the household tasks, the ageing population, the care of individuals with physical impairments and those in rehabilitation therapy [2]. It is very important for these systems to have good human-robot interaction (HRI) interfaces, allowing them to be easily accepted, usable and controllable by humans. In this field, various and different HRI issues arise such as making user interfaces that reduce the cognitive load of the robot interlocutor and allowing him to interact naturally and efficiently with it. Both verbal and non-verbal communications are necessary to establish an engaging interaction. The robot should be able to communicate with the user through both verbal and non-verbal channels, which requires technologies capable of being commanded through natural communication (e.g., speech and natural language, hand gestures, facial expressions), of fetching items, and of assisting with daily activities (e.g., dressing, feeding, moving independently).

In this paper, we present the research and development that are being done in the

Martins C., Teixeira A., Pedrosa E. and Almeida N..
Spoken Communication with CAMBADA@Home Service Robot.
DOI: 10.5220/0003878900270036
In *Proceedings of the 2nd International Living Usability Lab Workshop on AAL Latest Solutions, Trends and Applications* (AAL-2012), pages 27-36
ISBN: 978-989-8425-93-5

area of service robots by the CAMBADA@Home team, a research group from the Aveiro University and comprising students from the Department of Electronics, Telecommunications and Informatics and researchers from the Institute of Electronics and Telematics Engineering of Aveiro (IEETA).

In the scope of IEETA Transverse Activity on Intelligent Robotics, the project CAMBADA[1]@Home was created in January 2011 following the team past research done in the CARL[2] [3] project[3] and the experience in the CAMBADA [4] robotic soccer team. The objective of CARL project was to study the interrelations and integration of various dimensions of the problem of building an intelligent robot: human-robot interaction, sensory-motor skills and perception, decision-making capabilities and learning. The development of the CAMBADA soccer team started in 2003 and has participated in several national and international competitions, including RoboCup world championships (1st in 2008, 3rd in 2009, 2010 and 2011).

The CAMBADA@Home project aims to address the special issues that arise when one needs to develop services and assistive robot technology with high relevance for personal domestic applications in daily life [5]. The development of such a robotic platform for elderly care is part of a broader project named Living Usability Lab for Next Generation Networks[4], a collaborative effort between the industry and the academy that aims to develop and test technologies and services that give elderly people a quality lifestyle in their own homes while remaining active and independent.

According to these research lines, the team is working to participate in the RoboCup@Home competition [6], whose aim is to enhance the development of fully autonomous robots capable of assisting humans in everyday life (e.g. personal robot assistant, guide robot for the blind, robot care for elderly people). In terms of HRI an aim of the competition is to foster natural interaction with the robot using speech and gesture commands since it cannot be touched during the competition. Hence, a robust speech recognition interface with an easy usable dialog system is strictly necessary and long-term goal.

The proposed spoken dialog system, as well as the implemented speech processing components and their integration within the CAMBADA robotic platform, are presented in this paper. In the next section, we give an overview of the main challenges and objectives that spoken interaction interfaces have to overcome on service robotic platforms. The robot prototype that currently has our spoken dialog system running is briefly introduced in section 3, being our approach to implement a dynamic dialog system explained in section 4. Furthermore, in section 5 we present some application scenarios defined for RoboCup@Home competition, drawing some conclusions and future work guidelines in section 6.

[1] CAMBADA is an acronym of Cooperative Autonomous Mobile roBots with Advanced Distributed Architecture.

[2] CARL is an acronym of Communication, Action, Reasoning and Learning in Robotics.

[3] http://www.ieeta.pt/carl/

[4] http://www.livinglab.pt/

2 Challenges

As motivated in the introduction, a mobile service robot should be enabled to interact with humans in home environments in a natural way, and speech is one of the most intuitively ways for HRI. Moreover, if speech is used for interaction, it should be as simple and effective as natural communication between humans. This means that a spoken dialog interface should be able to deal with complex and rich information exchange scenarios. As such, natural language interaction is a challenging problem, not only because it requires sophisticated natural speech processing units (speech recognition, speech synthesis and language understanding/generation), but also because it raises issues such as robustness, mixed-initiative dialog, multimodal interaction, and cognitive modeling [7].

A dialog system has to be easily usable by humans that do not know the robot and has to prevent deadlocks in situations where the robot does not understand the human. Therefore, a spoken dialog between a human and a robot has to be managed somehow. The responsible component for that control is called a dialog manager. Building a dialog manager however is a challenging issue. There are principles for standard dialog handling [8], which have to be kept according to the target application. Some of these principles include features such as error handling, timing and turn taking, mixed-initiative interaction, confirmation of irreversible actions, emergency exit, alternatives processing, helping system, dialog flow transparency, etc.

One of the most important and persistent problems in the development of HRI interfaces is their lack of environmental robustness when confronted with understanding errors. Most of these errors stem from limitations in speech recognition technology [8]. In the context of spoken language interfaces, the accuracy is mainly affected by the amount and type of environmental noise, acoustic echo, variations in speaking styles (e.g. accents, native and non-native speakers), and various spontaneous speech phenomena such as disfluencies, hesitations, filled pauses, and so on.

In general, two different approaches can be choosen to increase the overall robustness of spoken dialog interfaces. One approach is to increase the accuracy of the speech recognition process. The other is to assume those errors and create the mechanisms for recovering from them through conversation, i.e. improving the dialog management process. This last approach is mainly followed when using generic and already implemented speech recognition systems (e.g. commercial systems), being necessary to rely on robust dialog managers.

3 Robotic Platform

The CAMBADA@Home robotic platform (see figure 1) is based on the CAMBADA robotic soccer platform [4]. The platform has a conical base with radius of 24 cm and height of 80 cm. The physical structure is built on a modular approach with three main layers. The top layer has the robot vision system that uses a low-cost sensor, the Microsoft Kinnect depth camera, and the speech input system, the Microsoft Kinnect microphone array. The middle layer houses the processing units, which collects data from the sensors and computes the commands to the actuators. Finally, the lowest layer is composed of the robot motion system. Since the project is still in its infancy, the platform does not include a robotic arm for the moment. The software architecture follows a distributed approach, with five control processes being executed concurrently by the robot's processing unit in Linux.

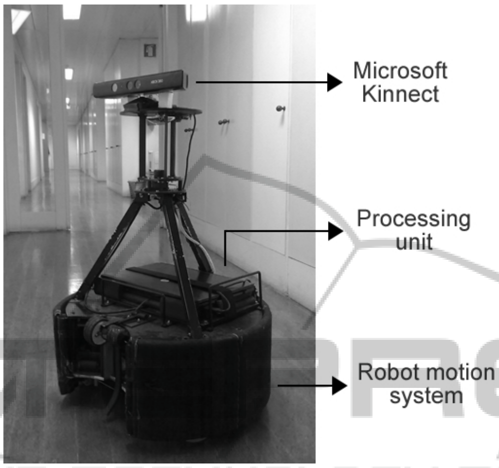

Fig. 1. CAMBADA@Home robotic platform.

An overview of the technical details can be found in [5]. As the natural interaction feature is the main topic of this paper, we will continue with a detailed description of the spoken dialog framework proposed.

4 A Spoken Dialog Framework for CAMBADA@Home

We have integrated some interaction facilities in our mobile service robot by means of a spoken dialog framework. The requirements for this speech-base interaction system resulted from the rulebook of the RoboCup@Home competition. According to its stated use-cases, we defined the following requirements:

- the speech recognition component should be speaker independent, have a small vocabulary and be robust against stationary and non-stationary environmental noise;
- the speech output should be intelligible and sound natural (at a distance of about 2 meters);
- the dialog manager system should be mixed-initiative, allowing both robot and user to start the action, provide or ask for help if no input is received or incorrect action is recognized (error handling), and ask for confirmation in case of irreversible actions.

Based on these requirements, we created the CAMBADA@Home spoken dialog system, whose architecture is shown in figure 2.

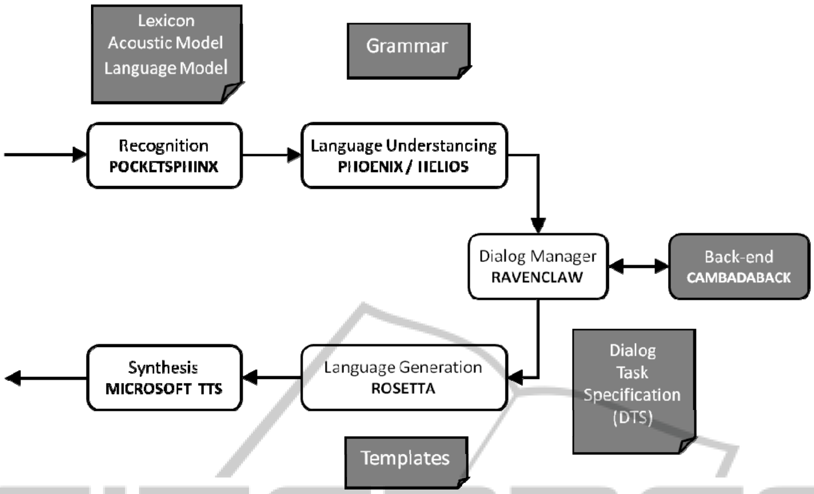

Fig. 2. CAMBADA@Home spoken dialog framework using RavenClaw/Olympus architecture.

Our system was built using RavenClaw/Olympus [9], an open-source framework that facilitates research and developments in task oriented conversational spoken language interfaces. Olympus [10] is an architecture for spoken dialog systems created at Carnegie Mellon University (CMU) and consisting of various components for recognition, natural language understanding, dialog management, natural language generation, speech synthesis, etc., and the corresponding communication infrastructure. In this framework, the RavenClaw, a task-independent dialogue engine, handles the dialogue management. Olympus has been used to develop various other systems that span different domains [11]; [12]; [13].

The clear separation between the domain independent components (dealing with conversational skills such as misunderstandings, the accuracy, repeats, focus shifts, etc.) versus the domain dependent components (such as the lexicon, the acoustic and the language models, the grammars for natural language understanding/generation and, mainly, the dialog task specifications (DTS)), is one of the main characteristics of RavenClaw/Olympus framework. That characteristic was one of the reasons why we chose this framework, since it allowed us, in a first approach, to focus our research effort on the development of the domain dependent resources of our system. Another reason that contributed to our decision of using Olympus architecture was the fact that being an open source framework, it assembles and provides all the necessary components for a quick development of a spoken dialog system, allowing us to easily plugging in our own modules into the architecture at any moment.

At the moment of writing, the main components implemented in our spoken dialog framework (fig. 2) include: a speech recognizer from CMU (PocketSphinx), a semantic parser (Phoenix), a dialog manager (RavenClaw), a natural language generator (Rosetta), a speech synthesis system (from Microsoft) and a back-end module (CAMBADABACK) to process the intercommunication between the framework and the robot processing unit. On the following sections, we briefly describe these components and their corresponding resources (the task-dependent components).

4.1 Speech Recognizer

In terms of hardware, two types of robot-mounted input systems are being tested: a directional microphone and a Microsoft Kinnect microphone array with noise reduction and echo cancellation. To deal with the high amount of non-stationary background noise and background speech, a close-speech detection framework - an energy/power based voice activation detection (VAD) is applied in parallel to noise robust speech recognition techniques.

At the time of writing, speech recognition is accomplished with Pocketsphinx, an open source decoding engine [14] using generic acoustic models. The recognition engine uses a class-based trigram and a lexicon of 499 words (containing some of the most task-independent frequent words and the task-dependent words gathered from tasks presented at previous RoboCup@Home competitions and including names, items and locations. The pronunciations for those words were generated using CMU dictionary [15]. Additionally, we are testing speech recognition results obtained by using the Microsoft Speech Platform [16]. For this propose both speaker dependent and speaker independent profiles are being tested.

4.2 Natural Language Understanding

In this first approach, our system is using Phoenix [17], the default semantic parser in RavenClaw/Olympus framework, to extract concepts from the recognition results. Phoenix uses a semantic grammar assembled by concatenating a set of pre-defined domain-independent rules with a set of domain-specific rules authored by us and according to the tasks defined for our dialog system.

The set of parsed hypotheses is then passed to Helios [18], a confidence annotation component that uses features from different knowledge sources (e.g., recognition, understanding, dialog) to compute a confidence score, forwarding the hypothesis with the highest score to RavenClaw, the dialog manager.

4.3 Natural Language Generation

The semantic output of the dialog manager is sent to Rosetta [19], a template-based language generation component. Like the Phoenix grammar, the Rosetta templates are assembled by concatenating a set of pre-defined templates, with a set of templates manually authored by us taking into account the specific tasks of our system.

4.4 Speech Synthesis

For robot “speak back” interaction and user feedback, external speakers mounted on the robot platform are used. The speech synthesis component is implemented by means of a concatenative system for speech output. For that propose, we are using Microsoft Speech SDK and a Cepstral Text-To-Speech synthetic masculine voice (David voice). Cepstral voices have a native audio format of 16kHz, 16bit, PCM, mono, and support SSML, VoiceXML tags, and Microsoft(R) SAPI standards, which allows voice customization compatible with Olympus framework. Moreover, we are testing some adaptation features such as using information on the distance from robot to user to dynamically change the output volume and the TTS rate from normal to slower according to the user's age.

4.5 Dialog Manager

As already said, our dialog system has been built using RavenClaw, a plan-based and task-independent dialog management framework [9]. As the perception of the environment is uncertain and human users may not always react, or react unexpectedly, there are some important conversational skills being automatically supported and enforced by the RavenClaw dialog engine (error handling, timing and turn taking, help, repeat, cancel, quit, start-over, etc.).

Our dialog system covers a restricted domain only, which is specific for the tasks the robot needs to perform. In RavenClaw architecture, the domain-specific aspects are captured by the dialog task specification (DTS), a plan-based description for the expected interaction tasks. Each DTS consists of a tree of dialog agencies, dialog agents and concepts. Dialog agents are located at the terminal positions in the tree, and implement a dialog action according to their type (Inform, Expect, Request or Execute action). Dialog agencies are located at non-terminal positions, and their purpose is to control the execution of sub nodes. And concepts representing entity values are used to store results. For dialog task specification, RavenClaw provides a proper language, the RavenClaw Task Specification Language (RCTSL). In our spoken dialog framework, we are defining a specific DTS for each one of the RoboCup@Home competition use-cases. In section 5 we give a brief overview of the first DTS implemented in our system and corresponding to some of the application scenarios defined for RoboCup'12.

Finally, to integrate the spoken dialog framework here created with CAMBADA@Home robotic platform, a back-end component (CAMBADABACK) has been developed. This component rules the communication between the RavenClaw dialog manager and the software layer controlling the physical structure. The communication callbacks between both are triggered by asynchronous messages whose syntax is represented by a set of predefined XML grammar rules.

5 Examples from RoboCup@Home Scenarios

As stated before, the development of CAMBADA@Home robotic platform is being done to cope with the tasks defined for the RoboCup@Home competition. As such, and in a first approach, we started by implementing two of those tasks: the "Robot Inspection" and the "Follow Me" tasks.

In the "Robot Inspection" task, whose focus is articulation and speech synthesis, the robot has to autonomously approach a table with some persons behind it, introduce himself and leave the room nextafterwards. In figure 3, we present a portion of the dialog task tree for the "Robot Inspection" task. It comprises two dialog agencies ("cambada", the root node, and "move", the agency controlling the moving action nodes), four dialog agents, the ones controlling dialog actions, and one concept ("move") capturing the expected input for this task (an order from someone asking the robot to move or leave room). The dialog starts when CAMBADABACK receives a signal from the robot processing unit. At that moment, the inform agent "Presentation" is activated, and an introduction message is synthesized and send to the output. Then a request message is sent to the output by the request agent "AskMove", and the dialog system waits until the "move" concept is filled with an order from the operator telling the robot to leave or move. Received that message, the execution agent "OrderMove" sends an order to the robot processing unit through CAMBADABACK. Finally, the "Exit" agent sends a leaving message to the output and the dialog task ends.

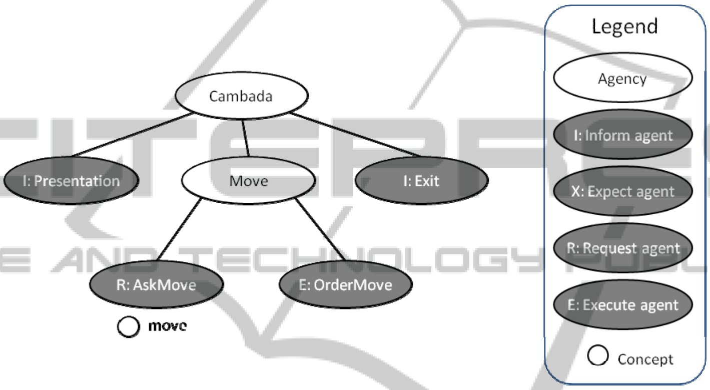

Fig. 3. A portion of the dialog task tree for the "Robot Inspection" task.

The "Follow Me" focuses on tracking and recognition of an unknown person. In this task, an operator goes to the robot and tells it to follow him. The robot starts a calibration process (to identify the operator), during which it gives the operator some instructions. After that, it announces the calibration is done and starts following him. Then the operator tells the robot to stop. The robot waits for 10 seconds and then start following him again. The operator orders the robot to stop again. After some time the operator and another unknown person walk towards the robot, and the operator asks the robot to go to him (go to owner). After recognizing the operator, the robot goes towards him. At the end of the task, if the operator congratulates the Robot, it thanks him and the task is considered finished. For this task, another DTS has been developed and integrated in our framework.

6 Conclusions and Outlook

In this paper, we presented a spoken dialog framework that dynamically processes speech dialogs for human-robot interaction in intelligent home environments. The created framework was integrated on the mobile service robot CAMBADA@Home. This robotic platform is being used to implement and test several household related scenarios as specified by the RoboCup@Home competition that aims to develop service and assistive robot technology with high relevance for future domestic applications.

At the time of writing, the core components of our spoken dialog framework were integrated, allowing us to run and test some of the RoboCup@Home tasks. As additional features become available on the CAMBADA@Home robotic platform, more home-centered tasks foreseen in RoboCup@Home rulebook are planned to be added to its speech interface repository. This includes tasks related to object detection/recognition/manipulation, human detection/recognition, and other tasks with increased complexity in terms of speech recognition and dialog management, and where actions have to be carried out without predefined order.

As future work, we intend to investigate further more the issues related to the overall system accuracy and robustness, especially the ones related to the speech recognition component. Using Microsoft Kinnect microphone array as an input unit, we are exploring its audio features and processing tools to better integrate them in our spoken dialog framework. Still related to the speech recognition component, we are investigating various approaches by combining different types of acoustic and language models for restricted domains. Moreover, due to the multimodal communication features present in the robotic framework, we will pursuit research efforts to improve the overall dialog system performance by taking advantage of information from other system components such as localization and vision. The CAMBADA@Home team plans to build on these results in order to participate in the RoboCup@Home challenges at the RoboCup'2012.

Acknowledgements

We would like to thank people from the Olympus framework developer's team for providing it, and for helping and answering our technical questions. This work is part of the Living Usability Lab for Next Generation Networks (www.livinglab.pt) project, a QREN project, co-funded by COMPETE and FEDER.

References

1. EUROP Executive Board Committee: Robotic Visions (To 2020 and Beyond). The Strategic Research Agenda for Robotics in Europe, July 2009
2. Tapus, A., Mataric, M. and Scassellati, B.: The grand challenges in socially assistive robotics. IEEE Robotics and Automation Magazine, vol. 14, no. 1, March 2007
3. Lopes, L.: Carl: from Situated Activity to Language Level Interaction and Learning, Proc. IEEE Int'l Conf. on Intelligent Robots and Systems (IROS), Lausanne, Switzerland, p. 890-896, 2002
4. Neves, A., Azevedo, J., Cunha, B., Lau, N., Silva, J., Santos, F., Corrente, G., Martins, D., Figueiredo, N., Pereira, A., Almeida, L., Lopes, L., Pinho, A., Rodrigues, J. and Pedreiras, P.: CAMBADA soccer team - from robot architecture to multiagent coordination. In Robot Soccer. INTECH, January 2010
5. Cunha, J., Neves, A., Azevedo, J., Cunha, B., Lau, N. and Pereira, A.: A Mobile Robotic Platform for Elderly Care. AAL Workshop, BIOSTEC 2011, Rome, Italy, 2011

6. RoboCup@Home League main page [Online] [Visited on 15 Oct. 2011]. Available at http://www.robocupathome.org/
7. Goodrich, M. and Schultz, A.: Human–Robot Interaction - A Survey. Foundations and Trends R in Human–Computer Interaction, vol. 1, no. 3 (2007) 203–275
8. Breuer, T.: Advanced Speech-based HRI for a Mobile Manipulator. B-IT Master Studies Autonomous Systems, Bonn-Rhein-Sieg University of Applied Sciences, September 2008
9. Bohus, D. and Rudnicky, A.: The RavenClaw dialog management framework - Architecture and systems". Computer Speech and Language 23 (2009) 332-361
10. Bohus, D., Raux, A. Harris, T., Eskenazi, M. and Rudnicky, A.: Olympus - an open-source framework for conversational spoken language interface research. Bridging the Gap: Academic and Industrial Research in Dialog Technology Workshop at HLT/NAACL, 2007
11. Bohus, D., Grau, S., Huggins-Danes, D., Keri, V., Krishna, G., Kumar, R., Raux, A. and Tomko, A.: Conquest - an Open-Source Dialog System for Conferences. HLT-NAACL, 2007
12. Stenchikova, S., Mucha, B., Hoffman, S. and Stent, A.: RavenCalendar - A Multimodal Dialog System for Managing a Personal Calendar. HLT-NAACL, 2007
13. Harris, T., Banerjee, S., Rudnicky, A., Sison, J., Bodine, K., and Black, A.: A Research Platform for Multi-Agent Dialogue Dynamics. Proceedings of the IEEE International Workshop on Robotics and Human Interactive Communications, (2004) 497-502
14. Huang, X., Alleva, F., Hon, H.-W., Hwang, M.-Y., Lee, K.-F., Rosenfeld, R.: The SPHINX-II speech recognition system: an overview. Computer Speech and Language 7 (1992) 137–148
15. The CMU Pronouncing Dictionary [Online] [Visited on 15 Oct. 2011]. Available at http://www.speech.cs.cmu.edu/cgi-bin/cmudict
16. Microsoft Speech Platform [Online] [Visited on 15 Oct. 2011]. Available at http://msdn.microsoft.com/en-us/library/hh361572.aspx
17. Ward, W. and Issar, S.: Recent improvements in the CMU spoken language understanding system, in Proc. of the ARPA Human Language Technology Workshop, pages 213–216, Plainsboro, NJ, 1994
18. Bohus, D. and Rudnicky, A.: Integrating multiple knowlege sources for utterance-level confidence annotation in the CMU Communicator spoken dialog system. (Tech. Rep. No. CMU-CS-02-190). Pittsburgh, Pennsylvania: School of Computer Science, Carnegie Mellon University, 2002
19. Oh, A. H. and Rudnicky, A.: Stocastic language generation for spoken dialogue systems. Proceedings of the ANLP/NAACL workshop on conversational systems, (2000) 27-32