Cooperative Human Robot Interaction with the Nao Humanoid: Technical Description Paper for the Radical Dudes

Peter Ford Dominey[1], Stephane Lallee[1], Anne-lise Jouen[1], Maxime Petit[1], Jean-David Boucher[1], Gregoire Pointeau[1], Pierrick Cheminade[1], Carol Madden[1], Alfredo Weitzenfeld[2]

[1] INSERM U846 Stem Cell and Brain Research Institut, Robot Cognition Laboratory, 18 ave Doyen Lepine, 69675 Bron, France
(peter.dominey, stephane.lallee, anne-lise.jouen, maxime.petit, jean-david.boucher)@inserm.fr,
[2] Information Technology and Engineering Depts, University of South Florida - Polytechnic, 3433 Winter Lake Road
Lakeland, FL 33803
aweitzenfeld@ieee.org

Abstract. Humanoid robots will increasingly interact with humans in the everyday, at home level. Here we describe the physical and software configuration of our humanoid platform which is being developed in this context of human – robot cooperation. The research strategy is to use high quality commercially available robot platforms, and software for sensory-motor control, vision, and spoken language processing, in order to provide a high performance baseline system. From this baseline we then implement human-robot cooperation capabilities inspired by contemporary results in human cognitive development research. Because much of the technology that we use is off-the-shelf it is robust, versatile and reusable.

Keywords: humanoid robot, vision, sensory-motor control, spoken language, shared plans, cooperation

1 Introduction

The Radical Dudes team represents our fourth generation of participation in the RoboCup@home league. We initially participated as the EK-Knights in Bremen in 2006 and qualified for the finals, then as the Robot Cognition Laboratory in Atlanta 2007 where we also qualified for the finals. In both of those competitions we used the Sony Aibo ERS-7 platform, running the Urbi system, along with a 6DOF arm in 2007. In both cases our Open Challenge task involved the robot learning from the human via demonstration and spoken language. We continue to develop this aspect of human-robot cooperation. While the Aibo provided a robust sensory-motor platform, it was lacking in the ability to manipulate and co-manipulate objects in the context of human-robot interaction. This motivates our current platform choice for the Aldebaran Nao humanoid. Our first effort with the Nao was in Singapore in 2010 with an 11th out of 23 in phase 1, and 9 of 13 in phase 2. The research and platform development described here is part of our long term effort to apply principals of computational cognitive neuroscience to robot perception (Dominey & Boucher 2005) and human-robot cooperation (Dominey et al. 2007a,b; 2008, Weitzenfeld & Dominey 2006).

2 Hardware Description

The Aldebaran Nao is a 25 DOF humanoid robot. It is a medium size (57cm) entertainment robot that includes an onboard computer and networking capabilities at its core. Its open, programmable and evolving platform can handle multiple applications and is currently the most evolved humanoid robot available on the market. The onboard processor can run the YARP server (described below) and can be accessed via telnet connection over the internet via WIFI.

Fig. 1. The Aldebaran Nao. In this image, a spoken language command processed by RAD and sent over the internet was issued to activate a pre-recorded Choregraphe script for grasping an object and then standing up in preparation to carry the object to the user. This will contribute to the fetch and carry task.

.

More specifically, the Nao is equipped with the following: CPU: x86 AMD Geode with 500 MHz, Memory: 256 MB SDRAM and 1 Gb Flash memory, WiFi

(802.11g) and Ethernet, 2 x 640x480 camera with up to 30 frames per second (one pointing at the feet, and one point forward), Inertial measurement unit (2 gyro meters and 3 accelerometers), 2 bumper sensors and 2 ultrasonic distance sensors

3 Software Description

The robot is running on a Linux platform and can be programmed using a proprietary SDK called NaoQi. It supplies binding for languages including C, C++i. It is also compatible with the robot simulator Webots by Cyberbotics. Both NaoQi and Webots are available for Linux, Mac OS X and Windows.

In the context of the software architecture, our objective is to exploit as much as possible existing high performance software tools, specialized for different specific aspects of robot perception and cognition. The value added component that we provide is the framework for human-robot cooperation which holds these components together.

3.1 Behavior Editing With Choregraphe

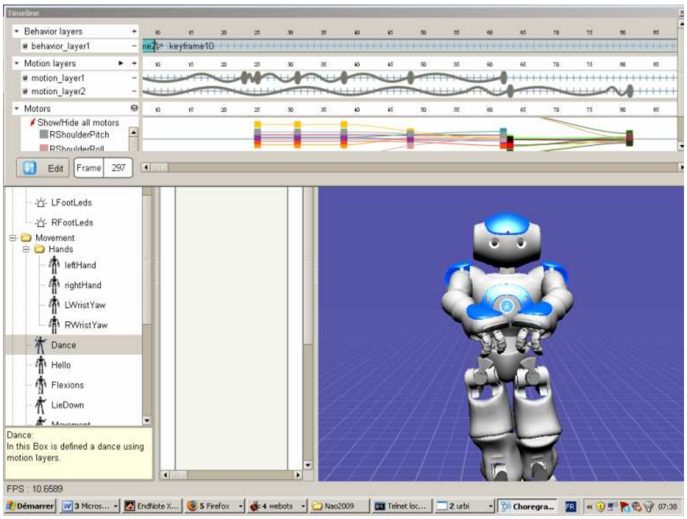

Fig. 2. Screenshot of the Choregraphe software GUI. Here we display the final posture of the Nao after testing and execution of the full body grasp behavior, corresponding to the actual robot configuration in Figure 1. Timeline in upper panel allows editing control of temporal sequences and transitions between predefined postures to create temporal behaviors.

Choregraphe is an intuitive GUI based tool that allows the user to rapidly create motion sequences, and to compose these together into complex, conditional behaviors. In particular, in a passive mode the user can physically manipulate the Nao into a desired configuration, selectively save the concerned joints, and proceed, and then link these postures together in a temporal sequence. These posture sequences can then be loaded onto the Nao via wifi, and they can then be called from the controller software.

3.3 Spikenet Vision

Vision is one of the most important capabilities in human-robot interaction. A strong vision capability will allow the robot to see and orient towards objects, to walk to them, grasp them and bring them to the user. It will allow the user to introduce new objects, new faces, new scenes to the robot. We have successfully used the SNV Spikenet Vision system (http://www.spikenet-technology.com/).

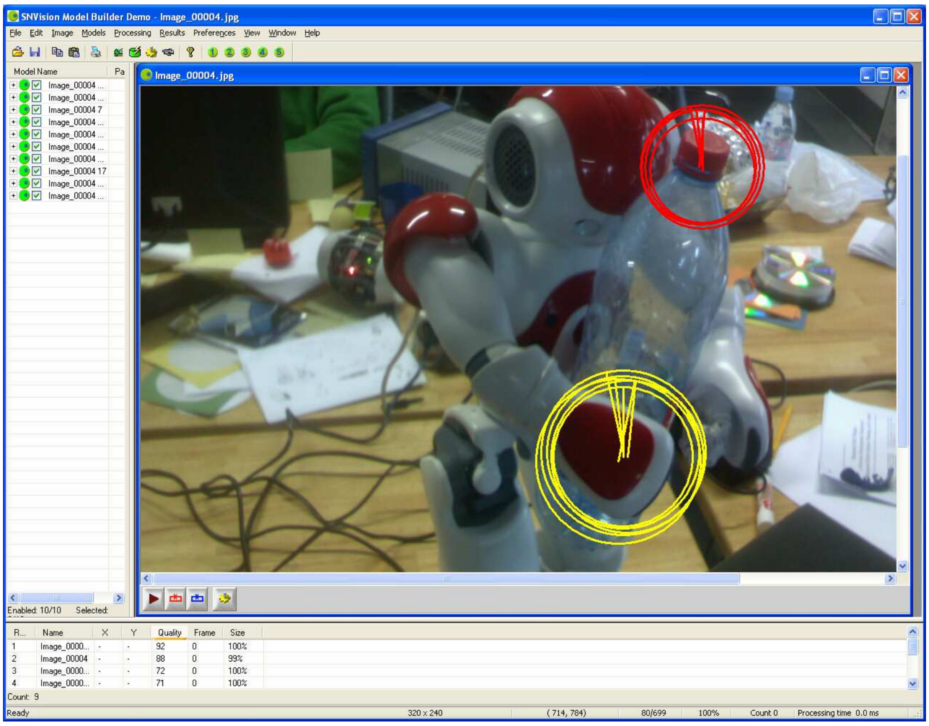

Fig. 3. Screenshot of the Spikenet Model Builder in recognition mode. Models have been constructed for recognition of the bottle top, and the hand of the robot. Here the recognition results are displayed as the superposition of oriented circles corresponding to the model recognition results on a frame of the camera image.

Recognition results from the Spikenet API are streamed and processed by a utility that associates different model files with their corresponding objects. This capability will be used in several of the @home tasks including fetch and carry, and fast follow.

3.4 RAD Spoken Language Processing

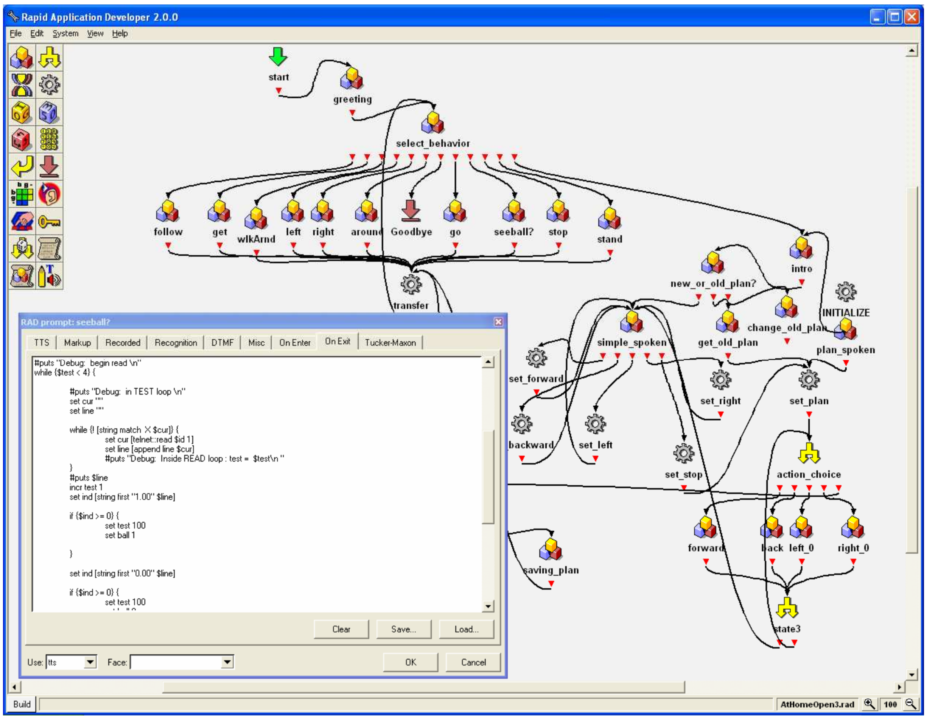

Fig. 4. Screenshot of the RAD program used in the 2006 open challenge. Nodes indicate states, with transitions conditional on spoken language recognition results and logical evaluation. Inset represents tcl code executed in one of these conditional state transitions.

At the core of the system, bringing these different software components together is the CSLU RAD Toolkit (http://cslu.cse.ogi.edu/toolkit/index.html). The toolkit provides spoken language recognition and synthesis. It is presented in a state based configuration in which the user can drag and drop state modules from a panel. These state module provide functions including conditional transition to new states based on the words and sentences recognized, and thus conditional execution of code based on current state, and spoken input from the user, as illustrated in Figure 4. In this example, code is executed which via a telnet connection to the robot, uses the Urbi interface to interrogate the robot vision system to determine if the red ball is visible.

3.4 Integrated System

The Nao and the above described software components are configured together as illustrated in Figure 5, exploiting the yarp communication framework (http://eris.liralab.it/yarp/).

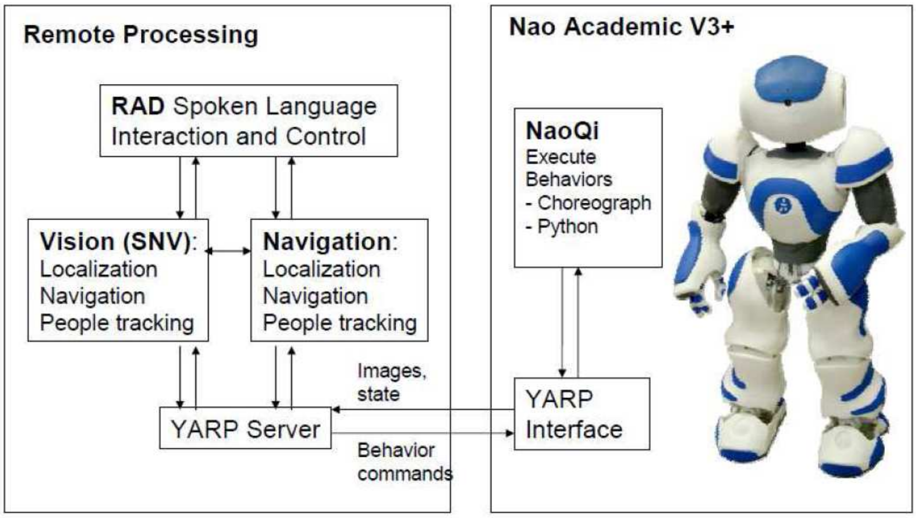

Fig. 5. Integrated system architecture overview. Running local on the Nao is the NaoQi system interfaced via Yarp.

On the Nao the NaoQi environment allows the use of behaviors that have been created and uploaded to the robot via Rad provides the backbone that connects the other components.

4 Innovative Technology and Scientific Contribution

The Nao is a state of the art humanoid platform. The Choregraphe system allows real-time behavior creation through physical and GUI interfaces. Urbi then allows these behaviors to be executed, along with others, based on sensory motor contingencies. It allows the creation of state machines for specification of robust sensorymotor behaviors. The RAD system provides a spoken language processing layer on top of this sensorimotor system.

4.1 Research Focus

Our research is focused on human – robot cooperation. Recent studies of human and primate cooperative behavior (Warneken et al. 2006) have revealed that helping requires that the agent have knowledge of the goal of the other agent, and motivation to act with and / or on behalf of that agent. Going farther, cooperation requires that the agent not only represent the goal of the other agent, but that she represent their shared goal, and that from this shared goal can derive a shared plan in which each agent knows both what she and her partner will do (Tomasello et al. 2006). Based on these studies we have begun to develop robotic implementations of these principals (Dominey 2006, Dominey & Warneken in press). Vision allows the robot to recognize action, and then form a representation of a shared plan in which the robot and the user each take their respective roles. Indeed given such a plan, the robot and human can exchange roles, and / or help each other.

4.2 Re-Usability by Other Groups

Because of the off-the-shelf nature of the major components of our system, the system itself and the usage of the individual components is well suited for reuse by other teams. Indeed, we are demonstrating a form of reuse already by applying much of what was done with the Aibo directly to the Nao. The platform is also well suited for the preliminary development of real world applications. In particular we focus on the concept of Ambient Assisted Living (http://www.aal-europe.eu/aal-2009-2) in which such robots will provide a social interface to aged people in their place of residence. The capabilities developed in the context of Robocup@Home have direct application in this new context.

Acknowledgements:

This technical work is supported by the EU FP7 ICT projects CHRIS, Organic and EFAA, and the French ANR projects Amorces and Comprendre.

References

Dominey PF (2005) Toward a construction-based account of shared intentions in social cognition. Comment on Tomasello et al. 2005, *Beh Brain Sci*. 28:5, p. 696.

Dominey PF, Alvarez M, Gao B, Jeambrun M, Weitzenfeld A, Medrano A (2005) Robot Command, Interrogation and Teaching via Social Interaction, *Proc. IEEE Conf. On Humanoid Robotics 2005*.

Dominey PF, Boucher (2005) Learning To Talk About Events From Narrated Video in the Construction Grammar Framework, *Artificial Intelligence*, 167 (2005) 31–61

Tomasello M, Carpenter M, Call J, Behne T, Moll HY (2005) Understanding and sharing intentions: The origins of cultural cognition, *Beh. Brain Sc*;. 28; 675-735.

Warneken F, Chen F, Tomasello M (2006) Cooperative Activities in Young Children and Chimpanzees, *Child Development*, 77(3) 640-663.

Warneken F, Tomasello M (2006) Altruistic helping in human infants and young chimpanzees, *Science*, 311, 1301-1303

Weitzenfeld A and Dominey PF, Cognitive Robotics: Command, Interrogation and Teaching in Robot Coaching, RoboCup Symposium 2006, June 19-20, Bremen, Germany.