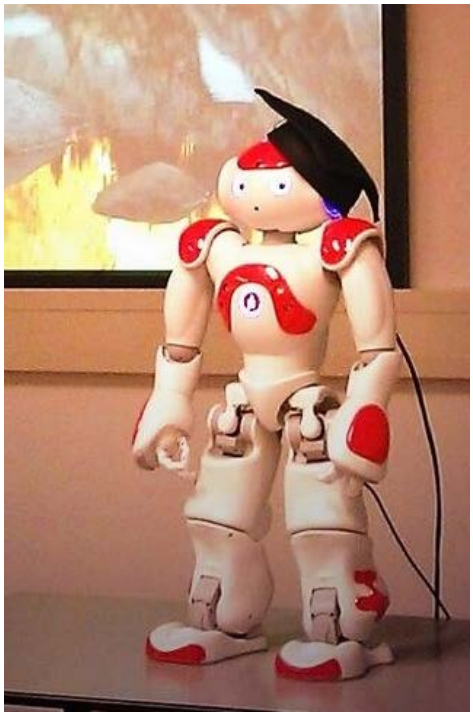
Question-Answering and Camera-Based Audience Analysis for the RoboTutor

Imagine a teacher that is able to answer all questions and who is fully aware of the audience that is being lectured. By augmenting the tools of a teacher using state-of-the-art software for question-answering and computer vision we can imitate some of the cognitive capabilities of human teachers. It is the aim of this project to make a significant contribution towards achieving this goal. Although the main focus of the RoboTutor project currently is on educational contexts, there is no reason to not think bigger; check out, for example, the Robot TED Talk Challenge: http://www.xprize.org/ted.

The starting point for the project is a robotic teacher called RoboTutor. The RoboTutor is a Nao robot that has been equipped with a script engine and special software that enables it to give a presentation to an audience. Instead of a virtual agent the RoboTutor is an embodied agent that presents a lecture (Nijholt, 2006). It can teach a lecture on robotics using a PowerPoint presentation and can pose quiz questions during the lecture using clickers. One of the advantages of a robot teaching robotics is that a robot teacher is able to *demonstrate* what it is teaching because it is a robot itself and it can perform the behaviors it is talking about. It can illustrate, for example, how many degrees of freedom the Nao has in its arm by moving its arm and it can show that it is able to view the audience by showing a picture of the audience that is taken by its camera on a presentation slide in real-time. See the appendix at the end for more details on the RoboTutor.



The goal of this project is twofold. First, the aim is to develop a **question-answering (Q&A) system** that is able to answer questions on robotics, using information that can be found on the Web. Second, the aim is to develop a camera-based **Audience Analysis System** that is able to analyze audience attention and to detect activities in the audience such as raising a finger to ask a question. The objective is to significantly extend the interaction capabilities of the current RoboTutor. Integrating these components into the overall system is part of the project work. This involves integrating the Q&A and audience analysis system into the control architecture of the RoboTutor. *Note that it is also possible to split the project into two subprojects and focus on only one of the two mentioned goals in your bachelor project.*

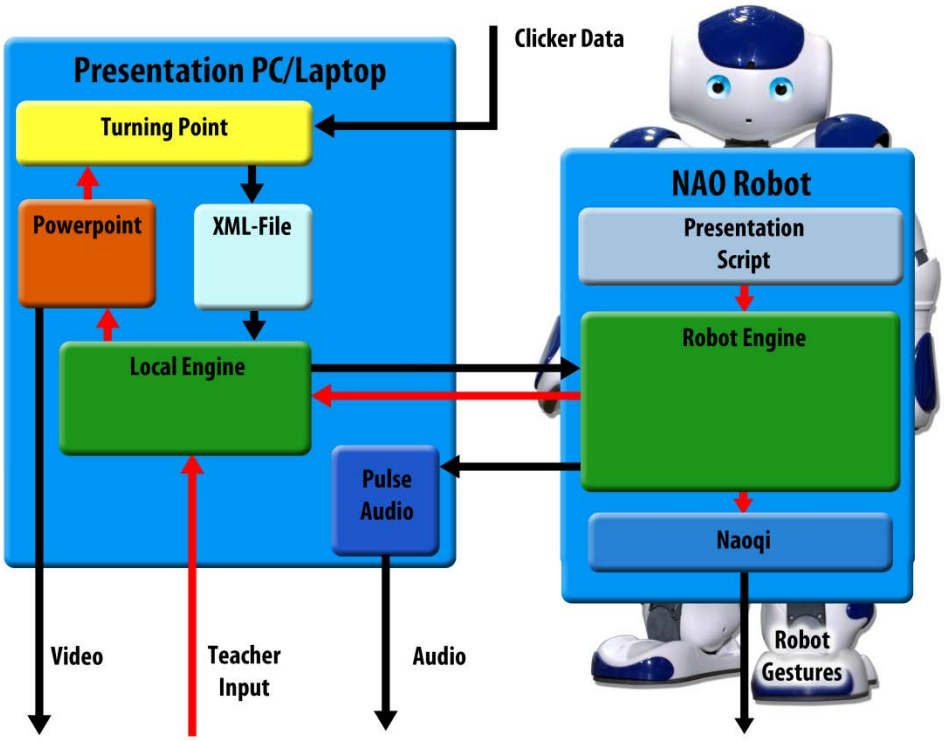


Figure 1 High-level Architecture of RoboTutor System

Question-Answering System for Audience Interaction

State-of-the-art question-answering technology that is available today is ready to be applied to real-time, real-world tasks (Kolomiyets & Moens, 2011). The famous Watson system developed by IBM that won Jeopardy! is an example of what can already be achieved today (Ferrucci, Levas, Bagchi, Gondek, & Mueller, 2013). The goal of this project is to develop a question-answering system that allows participants in the lecture audience to ask questions about robotics. In the project a QA system will be selected and implemented for the RoboTutor. The system should be able to connect to and search the Web for answers in real-time but should also maintain a database of answers to frequently asked questions (to be able function without access to the Web). The RoboTutor should provide an answer by using its built-in text-to-speech system but the answer should also be displayed on a slide in real-time. Participants in the audience should be able to use speech to ask questions. To this end, a microphone and speech-to-text system is available. Connecting and integrating the QA system to a speech recognition system and integrating it into the overall RoboTutor architecture (see *Figure 1*) is also part of the work to create a RoboTutor with Q&A functionality. The focus in this project will be on answering questions about robotics, but the system should be general enough such that it is possible in the future to also answer questions about other topics.

Camera-Based Audience Analysis System

State-of-the-art image processing technology that is available today is ready to be applied to real-time tasks such as tracking the faces of individuals in an audience such as a lecture room (Yu and Nakamura, 2010). The goal of this project is to develop an audience analysis system that allows the robot to more actively engage in interaction with participants in the lecture room. In the project a Camera-Based Computer Vision (CCV) system will be developed for the RoboTutor. The goals of

developing the system should make the robot appear more aware of its audience in various ways. It should, for example, show increased awareness of its audience by being able to look at participants and by directing its gaze in the right direction. This increased situation awareness should not be limited to detecting participants in the audience but also allow the system to detect activities performed by participants in the audience (e.g., identifying a student who raises finger to ask question). Finally, the system may be used to detect more general features of the audience related to, for example, a general level of attention. This can be used by the robot to pause for some time to (re-)capture audience attention, but may also be used to select and ask someone from the audience who pays attention to answer a question. Connecting and integrating the CCV system into the overall RoboTutor architecture (see *Figure 1*) is also part of this project Connecting and integrating the QA system to a speech recognition system and integrating it into the overall RoboTutor architecture (see *Figure 1*) is also part of the work to create a RoboTutor with Q&A functionality. Connecting and integrating the QA system to a speech recognition system and integrating it into the overall RoboTutor architecture (see *Figure 1*) is also part of the work to create a RoboTutor with Q&A functionality..

Overall System Integration

The overall goal would be to integrate and combine the capabilities of the Q&A and Audience Analysis Systems. The challenge is to combine these systems and make it possible for the RoboTutor to look at the participant while answering the question without any intervention of a human. The RoboTutor should be able to detect that a participant in the audience wants to ask a question, process the question posed by analyzing speech, find an answer, and present this to the audience again.

The Team Needed for the Job

Preferably, the team for this project:

- Has experience with programming in C++, C#, and/or Java.
- Is ambitious and looking for a challenge.
- Designs a system with the user in mind.

Work space is available in our RobotLab which is the base location for our Nao RoboTutor. During the project you will be working in this Robotlab that is part of the INSYGTLab on the 2nd floor of the EWI building.

Appendix: Software Architecture of the RoboTutor

In this appendix we briefly discuss the architecture of the current version of the RoboTutor (*Figure 1*). The architecture consists of various components:

- The **script engine** component: controls the behavior of the RoboTutor by executing a script.
- The **event-based** component: allows robot to respond to events, triggered by other systems.
- The **turning point** component: support for answering quiz questions using clickers.
- The **PowerPoint** component: support for controlling a PowerPoint slide presentation.
- The **mood expression** component: supports expressing various moods while presenting.
- **Repository of behaviors**: the robot has a repertoire of about 40 body movements it can perform and in addition can perform a dance and has various other capabilities.

The RoboTutor uses a **script engine** that runs on the Nao to present a lecture. A script consists of a series of commands. These commands allow the robot to choose ("branch") different courses of action based on the input received during a presentation, to change slides in a slideshow, and to perform other behaviors that the Nao is capable of. The script engine provides support for synchronizing and controlling non-verbal behavior (the "body language", gestures, etc.) and the verbal output of the robot (spoken text). Available behaviors include various behaviors for moving the arm, legs and head together or separately as a (human) presenter would (e.g., for performing a pointing behavior). Other behaviors such as dance behaviors are easily integrated by means of Nao's Choregraphe system. Other commands that are available include taking a picture of the audience, controlling the pace of the slideshow and pausing/resuming the execution of the script.

The RoboTutor is currently able to respond to a limited set of **events**. For example, a human (teacher) can pause and resume the execution of the script using a script GUI or by touching the head of the Nao. The RoboTutor can also respond to the answers that students give to a quiz question. The event mechanism needs to be extended with events such as a participant raising a finger to ask a question.

The Robotutor makes use of the **Turningpoint system** for presenting **PowerPoint** slides. This system consists of 'response cards' that are handed out to the audience. These cards are small electronic devices with a number of buttons. These cards can be used to answer multiple choice questions. In the presentation, the robot will show a slide with a question and asks the audience to respond. The results can then be shown on the screen and the robot will reply accordingly.

The **mood component** provides support for varying the mood expressed in body language (the behaviors of the robot). The robot can be made to appear more or less positive and energetic. A **repository** of 45 behaviors for which mood has been modulated is available.