

Human-Robot Dynamic Social Interaction

NTT9904-01

Progress Report: July 1, 2002—December 31, 2002

Rodney Brooks

Project Overview

NTT researchers are interested in the question of whether a physical robot produces a more direct emotional coupling with human beings than does a computer generated graphical image of a similar robot. At MIT we are building a robot that has human-like facial expressions and shoulder and neck gestures, and that perceives human motion and facial expressions. This is coupled to an emotional system so that the person and the robot naturally follow normal human communication social dynamics. This robot will be installed at the NTT Communications Science Laboratories in Kyoto where the response of human subjects will be measured and compared to their response to a graphical interface.

Progress Through December 2002

In 2000 and 2001 we delivered a preliminary robot to NTT, updated the design of the ultimate robot, called Kismet, resolved complex mechanical issues surrounding Kismet and fabricated Kismet components. Software infrastructure for Kismet was developed. In the first half of 2002 Kismet was completely assembled and implemented then transferred to CSL in Kyoto. Handoff was facilitated by a visit by two NTT CSL researchers: Kazuhiko Shinozawa and Futoshi Naya in June for 2 weeks. The NTT researchers communicated a description of their experimental methodology and current progress. MIT researchers teamed to familiarize the NTT researchers with trouble shooting, mechanical operations and software control of Kismet.

In the latter half of 2002, NTT prepared Kismet for experimental use. Kismet was fitted with a face shell that will play an integral role in the nature of social responses between Kismet and human subjects. MIT research has been to continue to develop software that can run on the NTT Kismet. We have used Cog as the experimental platform at MIT; NTT Kismet and Cog share the same software structure above their low level motor and frame grabbing code. The work at MIT in the second half of 2002 involved the development of an egocentric map.. An egocentric map keeps track of the locations of objects relative to the robot's body. The frame of reference for the objects in the world is centered within the robot and principally derived from its visual and manipulative field. This allows the robot to look away from one part of the scene, and still know what to expect when it looks back. This capability will be useful for NTT Kismet so that it can have more sophisticated interactions with people, knowing where they are even when it looks away.

We have also endowed Cog with a speech processing system where acoustic input can either come from a microphone array on the body of the robot, or a portable microphone held by a human. This is a practical arrangement since it means that under normal operation it is possible to speak to the robot by simply standing near it, but if the recognition rate degrades due to a large amount of noise in the background, it is still possible to operate the robot with the portable microphone. To further improve recognition performance, a low vocabulary size is used, with provision to easily introduce new words. In practice, relatively few words are needed for a particular task, so this is a very reasonable trade-off.

Research Plan for the Next Six Months

The egocentric map and speech processing modules will be combined with an object recognition module where the robot learns to locate and identify objects that it has previously touched. Our goal is to allow the robot to learn how to perform a task through demonstration by taking over the task incrementally.

The MIT team will also support the experiments using Kismet at NTT CSL in Kyoto. We will provide updated hardware as it is need. We will visit NTT at a critical time in the experiment schedule to provide on-site hardware support for the robot.