

Casual Communication with Robots using Speech Recognition Module

Koaru Hirota, Fangyan Dong

Tokyo Institute of Technology
G3-49, 4259, Nagatsuta, Midori-ku, Yokohama 226-8502, Japan
E-mail: {hirota, tou}@hrt.dis.titech.ac.jp

The Mascot Robot System is developed as a part of the "Development Project for a Common Basis of Next-Generation Robots" sponsored by NEDO (New Energy and industrial technology Development Organization) that is Japan's largest public R&D management organization for promoting the development of advanced industrial, environmental, new energy and energy conservation technologies. The Speech Recognition Module is also provided for this project by NEC Corporation and NEDO.

The main purpose of the mascot robot system is to perform experiments demonstrating the functioning of the speech recognition module mounted on household robots. The system is implemented as a network of multi-robots freely connected by RT middleware (RTM). It consists of 5 robots, i.e., 4 fixed robots (placed on a TV, a darts game machine, an information terminal, and a mini-bar) and 1 mobile robot. Each of them includes an eye robot, a speech recognition module, and a notebook PC that controls the robot and the speech recognition module. These robots connect together with a server through the internet by RTM, thus constituting the Robot System. The Mascot Robot System's functioning is demonstrated in an ordinary living room, where casual communication between 5 robots and 4 human beings (1 host, 2 guests, and 1 walk-in) is conducted based on speech recognition and mentality expression of eye robots.

Experimentally, each of the 5 speech recognition modules achieves above 90% average recognition rates, the target rate being 80%.

Noise filtering experiments are also performed. Using TV noise, correct recognition of 100 words (20 words uttered by 5 speakers) within 3 noise levels (70, 60 and 50dB) is verified.

Correct recognition is also verified for dictionary switching experiments using 2 dictionaries.

The system also has adequate sound source direction. In experiments using speech coming from 12 directions, the correct direction is recognized with an error $0 \sim \pm 4$ [deg] ($+9$ [deg] only in the case of 90 [deg]).

Furthermore, it is confirmed that the remaining 3 or 4 robots work correctly even when one or two robots, such as the TV robot and mobile robot, brake down suddenly.

14 RT Components with sizes ranging from 10kB to 64kB are implemented using C++ language to freely connect the 5 robots to a server.

In addition, practical RT component applications are developed, i.e. an information obtaining module, a information mining module, an information recommendation module, and a user interface. In experiments conducted with 6 subjects, it confirmed that the system output and the evaluations of 6 subjects are highly correlated, and that the profiling precision is improved by utilizing the speech recognition results.

As an advanced application of the speech recognition module, a robust speech recognition system that integrates the recognition results of multiple speech recognition modules. The recognition rate by the robust speech recognition system has improved 8.4 point compared with that of single recognition module. It is also shown that introduction of speech recognition to user profiling has improved the quality of user profiling.

The experimental results are shown by DVD files.