# Team eR@sers 2016 in the @Home League Team Description Paper

Hiroyuki Okada[*1], Takayuki Nagai[*2], Tomoaki Nakamura[*2], Komei Sugiura[*3],

Naoto Iwahashi[*4] and Tetsunari Inamura[*5]

Address
*1 Tamagawa University, 6-1-1 Tamagawa-gakuen,Machida-shi, TOKYO 194-8610, JAPAN
*2 The University of Electro-Communications
*3 National Institute of Information and Communications Technology, JAPAN
*4 Okayama Prefectural University
*5 National Institute of Informatics, JAPAN

E-mail
h.okada@tamagawa.ac.jp

Abstract. Team eR@sers has taken part in RoboCup@Home since 2008. 2008 was the first year of the eR@sers. The eR@sers achieved a first place at the Japan Open Competition 2008, 2009, 2010, 2011, 2012 and first place RoboCup 2008 ,2010 and second place RoboCup 2009, 2012. We have imploved the ability of robots with various techniques, which are going to be applied to other robot systems or social IT systems. We introduce them and our latest research briefly in this description.

## 1. Team History

The Japanese Robot Team eR@sers(erasers) is the result of a joint effort of five Japanese research groups:

Tamagawa University:the group of the College of Enginerring at Tamagawa University of Tokyo in Japan that is involved in the world championship RoboCup competitions in Four-legged since 2005. At the RoboCup 2006 Bremen , the team, FC Twaves, got the best results(best 16) in the team that participated from Japan. The members at Tamagawa University are interested in a compliant human-machine interaction architecture that is based on the machine intention recognition of the human. This work is motivated by the desire to minimize the need for classical direct human-machine interface and communication.

The University of Electro-Communications:the group of the Department of Electronic Engineering, The University of Electro-Communications in Japan that takes a role in the visual processing system in this team. The main avenue of the research group is to pursuit the real human-like intelligence, including multimodal information processing. 2008 was the first year of the eR@sers. The eR@sers achieved a first place at the Japan Open Competition. And in wonderful, In RoboCup 2008 in Suzhou, China we got a first place. In 2009, the eR@sers achieved a first place at the Japan Open Competition. And in RoboCup 2009 in Graz, Austria we got a second place. In 2010, the eR@sers achieved a first place at the Japan Open Competition. And in RoboCup 2010 in Singapore we got a world championship agein. This paper presents the main development efforts of the team in 2013.

National Institute of Information and Communications Technology (NICT) and Okayama Prefectural University : the group of NICT in Japan that is involved in the research of the computational mechanism which enable robots to learn the communication by language and actions through natural interaction with human.

National Institute of Informatics : Based on the success of the preliminary challenge of the @Home Simulation in RoboCup Japan Open 2013 Tokyo, and the demonstration in the international RoboCup 2013 Eindhoven, the proposal of a new RoboCup @Home Simulation challenge has obtained the recognition of the international RoboCup committee and held the demo challenge in the RoboCup 2015 China in July 2015.

## 2. Software architecture

### 2.1 Framework

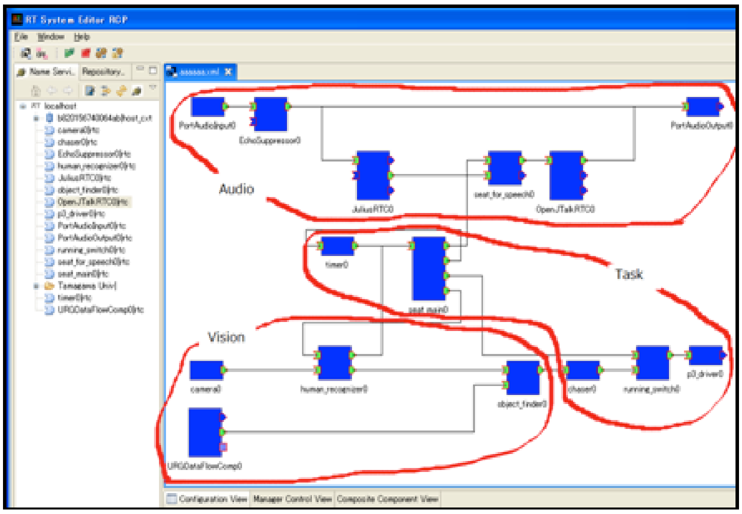


**Fig.** 1. eR@ser's software architecture

Fig.1. shows eR@sers software framework. In order to work on different parts of the robot system by different groups in different locations, we have used open source robotic technology middleware(OpenRTM) which is developed and distributed by Japan's National Institute of Advanced Industrial Science and

Technology. Both entire robots and various usable robotic functional elements are known as RT (Robot Technology). RT-Middleware is a platform to modularise and integrate a variety of robotic functional elements as software.

In RT-Middleware, an RT functional element is modularized as a software component called an RT-Component, and a robot is implemented by combining RTComponents on RT-Middleware. RT-Components have data ports and service ports to communicate with other components, and you can integrate various components easily by standardizing these interface specifications. All RT-Components have a common state machine inside, and you can manage a large number of components in higher-level application programs through integration. Moreover, since RTComponents have standard interfaces to alter internal parameters, you can reuse them without recompiling.

All components are connected through the "naming server" with GigE and have subscription information that describes required information for the processing in each component. The "Vision Module" is responsible for object/face detection and recognition tasks. The "Audio Module" provides speech recognition and text to speech functionalities. The "Task Module" works as a controller for each scenario of the @Home competitions. We have seven task modules in total, which are switched in accordance with the scenario.

2.2 Vision system.

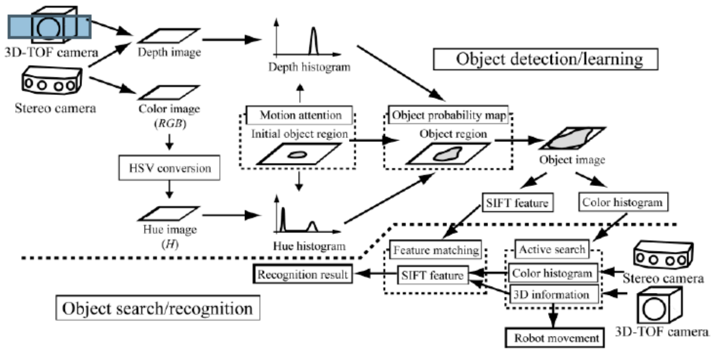


**Fig. 2.** Processing in the vision module

Object detection and learning. The vision system needs to extract objects from visual scene to learn objects in real time. A problem in extracting multiple objects from unlabeled images is that it is impossible to tell which part of the image corresponds to each individual object, and which part is irrelevant clutter. This system solves the problem using a priori knowledge that the segments with synchronous motion are parts of an identical object. Hence the motion detector is first employed in the object detection subsystem. As shown in Fig.2. the motion detector extracts the initial object region at first. Then,the object information such as color (hue) and depth is taken from the region. In particular, hue and depth histograms are taken from the region and normalized. Since these two histograms can be considered as probability density functions of the target object, the object probability map of each component at each pixel location can be easily obtained. The weighted sum of these two object probability maps results in the object probability map. The map is binarized followed by the connected component analysis, and then final object mask is obtained. In the learning phase, object images are simply collected, and then histogram and SIFT features are extracted. This information is used for the object search and recognition. It should be noted that the SIFT features are normalized using accurate 3D information in order to cope with affine distortion.

Object search/recognition. In order to recognize objects, we take two-step strategy. At first, the entire input image is scanned with calculating color histogram of a rectangular area in the image. Each histogram is collated with the histogram of the target object, and the area that gives the highest similarity is output as a position where the target object is detected. When the detected object is far away from the robot, the robot moves toward the object. When the object is close enough, a feature point matching is carried out in the detected region. We use the object matching algorithm that consists of two main stages: the selection of local scale invariant feature(SIFT), and the matching of constellations of such key points.

### 2.3 Speech Processing

Our speech technologies include noise robust speech recognition, high quality textto-speech conversion, and many-to-one voice conversion.

Speech Recognition. For speech recognition, HMM-based speech recognition software developed at National Institute of Information and Communications Technology(NICT) is used. Speech recognition system should be robust enough to recognize speech in noisy environments with various speaking styles. For acoustic modeling, the most important problem to solve is how to efficiently capture contextual and temporal variations in training speech and properly model them with fewer parameters. MDL-SSS creates speaker-independent models by data-driven clustering with contextual information based on minimul description length (MDL). This leads to high performance in large vocabularly continuous speech recognition. For front-end processing of speech recognition, adaptive noise reduction technique achieves the robustness of speech recognition against background noise. Gaussian mixure models for both speech and noise are used to form Wiener filter, and they are adapted according to input acoustic signal. To estimate the non-stationary noise sequences a particle filterbased sequential noise estimation method is used. In the proposed method, the particle filter is defined by a dynamical system based on Polyak averaging and feedback. A switching dynamical system is also applied into the particle filter to cope with the state transition characteristics of non-stationary noise. The method improves speech recognition accuracy even if noise is non-stationary.

Text-to-Speech Conversion. The concatenative text-to-speech conversion system, developed at NICT and named XIMIRA, is used. XIMERA is based on corpus-based technologies. The prominent features of XIMERA are as follows:

- large corpora (a 110-hours corpus of a Japanese male, a 60-hours corpus of a Japanese female, a 20-hours corpus of a Chinese female, and a 16-hours corpus of an English male)
- HMM-based generation of prosodic parameters;
- cost function for segment selection optimized based on perceptual experiments.

The result of evaluation test showed that XIMERA outperformed commercial TTS systems currently available in the market.

Voice Conversion. Voice conversion is a technology for converting a certain speaker's voice into another speaker's voice. Many-to-one voice conversion realizes the conversion from arbitrary user's voice as source to a target speaker's one. In our robot, arbitrary user's input voice is converted into robot's voice. Eigenvoice conversion is applied in the voice conversion method. Using multiple parallel data sets consisting of utterance pairs of the user and multiple pre-stored speakers, an eigenvoice Gaussian mixture model (EV-GMM) is trained in advance. Unsupervised adaptation of the EVGMM is available to construct the conversion model for arbitrary source speakers in many-to-one VC using only a small amount of their speech data.

## 3 RoboCup @Home Simulation

Research on high level human-robot interaction systems that aims skill acquisition, concept learning, modification of dialogue strategy and so on requires large-scaled experience database based on social and embodied interaction experiments. However, if we use real robot systems, costs for development of robots and performing many experiments will be too huge. If we choose virtual robot simulator, limitation arises on embodied interaction between virtual robots and real users. We thus propose an enhanced robot simulator that enables multiuser to connect to central simulation world, and enables users to join the virtual world through immersive user interface. As an example task, we propose an application to RoboCup@Home tasks.

Another function that can be simulated is using natural language and gesture instruction to recognize which object is being referred to by the user. A user might ask the robot to move and/or manipulate an object by saying something like, “Please bring that dish to the dining table” while pointing to the dish. If the pointing and/or speech are vague, the robot should be able to ask appropriate questions to remove the uncertainty. Such dialogue management is a high-level interaction function inherent in high-level HRI.

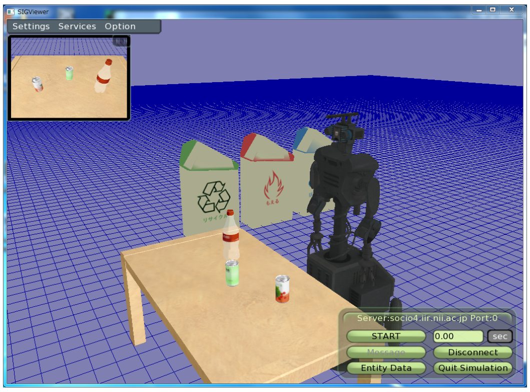

Fig3 Screen-shot of Clean Up task

**4 Hardware**

This year we take the robot TOYOTA HSR(Fig.4) to the competition.

4.1 Concept

The Human Support Robot (HSR) is being developed to assist people in their everyday activities. In the future, the HSR will coexist with family members in the home, providing support to improve living conditions and the overall quality of life.

4.2 Features

Three key features make it possible for the HSR to operate in indoor environments around people.

- Compact, Lightweight Body

To better accommodate a wide variety of households, the HSR needs to be lightweight and maneuverable. An articulated arm and telescoping body allow the HSR to cover a large workspace despite its compact footprint.

- Safe Interaction

Realizing that contact between human and robot is an essential aspect of support in domestic situations, a safety-conscious design was a top priority for the HSR. The robot's arm uses little power and moves slowly to prevent accidents and injuries. Obstacle avoidance and collision detection help the HSR to operate safely in a human-centric environment.

- Simple Interface

HSR can be controlled intuitively through voice commands or a simple graphical user interface via any number of common handheld touchscreen-enabled devices, such as tablet PCs and smartphones.

4.3 Functions

HSR has three basic modes: Pick-up, Fetch, and Manual Control.

- Pick-up
  The arm has a simple gripper to pick up objects such as pens and TV remotes, while thinner hard-to-grasp objects like paper or cards can be lifted off the floor using a small vacuum installed in the hand.

- Fetch
  Using voice commands or the touch-screen GUI, the user can command the robot to retrieve objects from boxes and shelves by simply specifying what to fetch.

- Manual Control
  Tasks that are currently beyond the scope of HSR's autonomous capability can be performed manually via the user interface. Manual control is also useful for remote operation ("telepresence"), which would allow caregivers and family members to communicate with the robot's owner over Skype or other services, by means of a display on top of its head.

4.4 Technology

- Folding Arm
  HSR is intended to help out around the home by fetching things, opening curtains, and picking up fallen objects. Along with a telescoping body, the robot's single arm can extend to pick stuff up from the floor or atop tables and high counters. When not in use, the arm is designed to fold in tightly to reduce its body's overall diameter.
- Flexible Hand
  Attached to the HSR's arm is a two-finger gripper, which is soft to the touch. This flexible hand conforms to the shape of objects it grasps, and includes a pressurized suction pad to lift thin items such as cards or paper.
- Object Recognition & Grasp Planning
  Object recognition algorithms allow HSR to understand the size and shape of items tasked to pick-up or grasp. This information is used to compute an appropriate path for the arm and position of the hand.
- Environment Recognition and Autonomous Mobility
  Onboard sensors keep HSR apprised of its surroundings, so that it can safely navigate inside the home, avoiding obstacles while continuing along the optimum route to its instructed destination.
- Remote Functions
  Family members and caregivers can access and operate HSR using a network-enabled device to perform the following tasks:
  - Remote Control
    Perform household tasks (Retrieve objects, open curtains, etc)
  - Remote Monitoring
    Watch over a disabled family member or check-in on an empty house
  - Remote Communication
    Video chat with family members ("telepresence")

4.5 Spec

- Body diameter 430mm
- Body height 1,005-1,350mm
- Weight 37kg
- Arm length 600mm
- Shoulder height 340-1,030mm
- Max payload 1.2kg
- Max object width 130mm
- Max speed 0.8km/h
- Max step size 5mm
- Max incline 5°

**Fig. 4.** HSR

## 5 The contents of the web site

Our relevant publications, technical reports, as well as videos and pictures are available in :

Official website: **https://sites.google.com/site/erasers2050/home/**

Photos and Videos of the robot:

**https://sites.google.com/site/erasers2050/photos-movies/**

Reference

1. Tetsunari Inamura, Jeffrey Too Chuan Tan, Komei Sugiura, Takayuki Nagai and Hiroyuki Okada: "Development of RoboCup@Home Simulation towards Long-term Large Scale HRI," Proc. of the RoboCup International Symposium 2013.
2. Noriaki ANDO, Shinji KURIHARA, Geoffrey BIGGS, Takeshi SAKAMOTO, Hiroyuki NAKAMOTO, "Software Deployment Infrastructure for Component Based RT-Systems", Journal of Robotics and Mechatronics, Vol.23, No.3, pp.350-359(2011)
3. Tomoaki Nakamura, Takayuki Nagai, and Naoto Iwahashi, "Multimodal Categorization by Hierarchical Dirichlet Process", IEEE/RSJ International Conference on Intelligent Robots and Systems, pp1520-1525, California, Sep.2011
4. Tomoaki Nakamura, Takayuki Nagai, and Naoto Iwahashi, "Bag of Multimodal LDA Models for Concept Formation", IEEE International Conference on Robotics and Automation, pp.6233-6238, Shanghai, May2011
5. Tomoaki Nakamura, Takaya Araki, Takayuki Nagai, Naoto Iwahashi, "Grounding of Word Meanings in LDA-Based Multimodal Concepts ", Advanced Robotics, 25, pp.2189-2206(2011)
6. Tomoaki Nakamura, Komei Sugiura, Takayuki Nagai, Naoto Iwahashi, Tomoki Toda, Hiroyuki Okada, and Takashi Omori, "Learning Novel Objects for Extended Mobile Manipulation", Journal of Intelligent and Robotic Systems, pp.1-18, Jul. (2011)