

Competition design to evaluate cognitive functions in human-robot interaction based on immersive VR

Tetsunari Inamura^{1,2} and Yoshiaki Mizuchi¹

¹ National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan,

`inamura[at]nii[dot]ac[dot]jp`,

² SOKENDAI(The Graduate University for Advanced Studies), 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan

Abstract. While RoboCup Soccer and RoboCup Rescue have simulation leagues, RoboCup@Home does not. One reason for the difficulty of creating a RoboCup@Home simulation is that robot users must be present. Almost all existing tasks in RoboCup@Home depend on communication between humans and robots. For human-robot interaction in a simulator, a user model or avatar should be designed and implemented. Furthermore, behavior of real humans often lead to unfair conditions between participant teams. Since the one-shot trial is the standard evaluation style in the current RoboCup, human behavior is quite difficult to evaluate from a statistical point of view. We propose a novel software platform for statistically evaluating human-robot interaction in competitions. With the help of cloud computing and an immersive VR system, cognitive and social human-robot interaction can be carried out and measured as objective data in a VR environment. In this paper, we explain the novel platform and propose two kinds of competition design with the aim of evaluating social and cognitive human-robot interaction from a statistical point of view.

Keywords: RoboCup@Home, Immersive Virtual Reality, Human-Robot Interaction, Evaluation of Social and Cognitive behavior

1 Introduction

We have been developing an immersive virtual reality(VR) system for a RoboCup@Home simulation[1]. While RoboCup Soccer and RoboCup Rescue have simulation leagues, RoboCup@Home does not. One reason for the difficulty of realizing a RoboCup@Home simulation is that robot users must be present. Almost all existing tasks in RoboCup@Home³ depend on communication between a robot and humans. For human-robot interaction in a simulator, a user model or avatar should be designed and implemented. Although this problem could be solved by

³ <http://www.robocupathome.org/rules>

using our immersive VR system, another difficulty with evaluating the quality of human-robot interaction still remains.

Not only RoboCup@Home but also RoboCup Soccer, Robocup Rescue, and other leagues tend to evaluate physical actions such as grasping, navigation, object tracking, and object/speech recognition because it is easy to evaluate them with objective sensor signals or ground truth data. However, evaluating the quality of human-robot interaction, such as the impressions of an individual user and whether a robot utterance is easy to understand, involve dealing with cognitive events, which are difficult to observe as objective sensor signals. One of the ultimate aims of RoboCup@Home is to realize intelligent personal robots that operate in daily life. While evaluating such social and cognitive functions is important, tasks for real competitions with time and space limitations cannot be designed for such evaluation. For example, using questionnaires is one conventional method for evaluating the social and cognitive functions of robots; however, this is difficult in real competitions because the number of samples that can be obtained is quite small.

Therefore, we propose a novel platform for competition design that can be used to evaluate the social and cognitive functions of intelligent robots through VR simulation. In Section 2, we propose a novel software platform that integrates ROS and Unity middleware to realize a seamless development environment for VR interaction between humans and robots. After that, we propose two tasks as examples of task design for evaluating social and cognitive functions. In Section 3, we propose the Interactive CleanUp task, which is aimed for statistically evaluating human-robot interaction. In Section 4, the Human Navigation task is proposed to observe and evaluate human behavior in terms of the quality of robot utterances. In Section 5, we discuss the feasibility of social and cognitive evaluation based on our proposed software platform.

2 Integrating Cloud-based immersive VR and ROS

Software for a RoboCup@Home simulation should enable both a conventional robot simulation and a real-time and immersive VR system for human-robot interaction to be integrated. Table 1 shows software systems and their pros and cons. Since conventional robot simulators focus mainly on simulating the physical world and modeling environments, the presence of human users is not discussed and implemented. There is currently no platform that enables the conventional robot simulation based on ROS middleware to be integrated with immersive VR applications for real-time human-robot interaction.

Our previous software platform, SIGVerse (ver.2)[2], was adopted for the first RoboCup@Home simulation trial in JapanOpen 2013 [1]. It enables users to log in to a VR avatar that makes it possible to communicate with virtual robots; however, it was constrained in that a variety of VR devices could not be used due to there being exclusive APIs. The API design was also too limited to fully support ROS middleware.

Table 1. Systems for robot simulation and VR interfaces; functions and limitations.

Platform	Graphic fidelity	Physics performance	Scalability of environment	Robotic middleware	Human immersion	Multi-client simulation	Licensing
Gazebo [5]	Very good	Very good	Good	ROS	Unsupported	Unsupported	Open source
USARSim [6]	Very good	Excellent	Very good	ROS	Unsupported	Supported	Commercial
V-REP [9]	Very good	Very good	Good	ROS	Unsupported	Supported	Open source
Choreonoid [7]	Very good	Very good	Not good	OpenRTM	Unsupported	Unsupported	Open source
Open-HRP [8]	Very good	Very good	Not good	OpenRTM	Unsupported	Unsupported	Open source
Webots [10]	Excellent	Very good	Very good	ROS, NaoQi	Unsupported	Supported	Commercial
SIGVerse(Ver. 2) [2]	Very good	Good	Good	None	Supported	Supported	Open source
SIGVerse(Ver. 3)	Excellent	Excellent	Excellent	ROS	Supported	Supported	Open source

Thus, we developed a new version of the SIGVerse platform that integrates Unity and ROS middleware to support a variety of VR devices and software resources created by the ROS community. Figure 1 shows the architecture of SIGVerse (ver. 3) in detail. SIGVerse is a server/client system. An identical VR environment (scene), composed of 3D object models, such as user avatars, robots, and furniture, is shared among servers and clients. Events in the server and clients are synchronized with Unity’s built-in networking technology.

Participants can log in to an avatar via a VR interface such as a head-mounted display (HMD), motion capture devices, and audio headsets. In accordance with the input from VR devices, the behavior of the participant is reflected in the avatar through scripts attached to objects. Perceptual information such as perspective visual feedback is provided to the participant. Thus, participants can interact with a virtual environment in a manner similar to a real environment.

The system has a bridging mechanism between the ROS and Unity middleware that is based on JSON, which is provided by a WebSocket, and BSON, which is provided by an exclusive TCP/IP connection. Several works have proposed connecting Unity and ROS middleware[3, 4]; however, huge data such as image data is difficult to send from Unity to ROS middleware in real-time. We measured the frequency of sending successive pairs of 900-[KB] RGB images and 600-[KB] depth image frames from a desktop computer, which had an Intel Xeon E5-2687W CPU, to a virtual machine on the same computer. Although the average frequency with JSON-based bridging was only 0.55 [fps], the frequency with BSON-based bridging was 57.60 [fps]. According to the specifications of standard 1-[Gbps] network interface cards, our platform has the potential to send four raw RGB images with a resolution of 640×480 captured by seven cameras within 30 [ms]. Software for controlling virtual robots can be used in real robots as well without any modification and vice versa.

Information for reproducing multimodal interaction experiences is stored on a cloud server as a large-scale dataset of embodied and social information. By

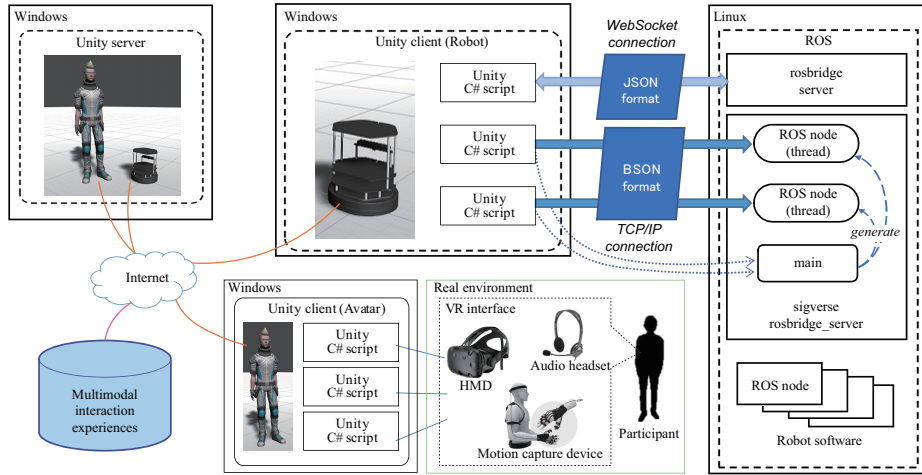


Fig. 1. Software architecture of SIGVerse

sharing such information, users can reproduce and analyze multimodal interactions after an experiment.

From the next section, we propose two tasks as examples of statistically evaluating social and cognitive functions on the basis of the VR software platform.

3 Task I: Interactive Cleanup

3.1 Task

Human behavior is one of the most important factors in the RoboCup@Home tasks. In the FollowMe task, the walking behavior of a real human is the target of recognition. In the Restaurant task, the instructions given by a human (referee) are observed by robots. A common limitation with using real human behavior is that behavior cannot be repeated several times to achieve a uniform condition for all task sessions. Thus, unfair conditions between each participant team are sometimes indicated to be a problem.

We propose a new task named Interactive Cleanup to solve the above problem in a VR environment. This task evaluates the ability of a robot to understand pointing gestures made by humans. The robot has to select a target object, which is pointed at by a human, grasp the object, and put it into a trash can. The target trash can is also pointed at with a pointing gesture. Therefore, the robot always has to observe the human’s gesture to understand the target object and target trash can. Figure 2 shows a picture of the VR task environment.

The procedure of the task is as follows.

- A human avatar points to one of the objects on a table. The human uses both right and left arms.



Fig. 2. VR environment for Interactive CleanUp

- After pointing to a target object, the human points to one of the trash cans.
- The robot should take the target object, move to the target trash can, and put the object into it.
- The procedure returns to step 1, and the task is repeated 36 times. A variety of pointing gestures made by the human avatar should be recorded beforehand. Different gestures should be reproduced in each session.

3.2 Evaluation result

This task was carried out at RoboCup JapanOpen 2016. Three teams took part in this task. The score distribution of the teams is shown in Fig. 3 and Table 2.

The vertical axes in Fig.3 indicate the score for each session; the horizontal axes correspond to the session index. The robot could get a point when it grasped

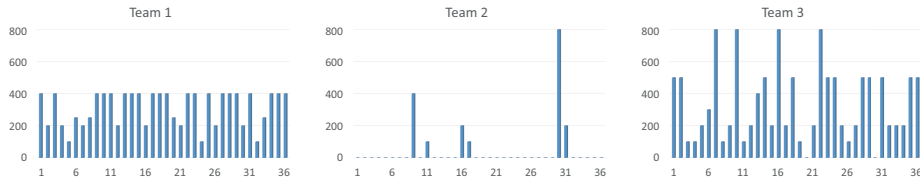


Fig. 3. Score distribution for Interactive CleanUp task

Table 2. Score stats for Interactive CleanUp task

Team	Average	Standard error
Team 1	314	108
Team 2	50	152
Team 3	339	234

the correct target or completely put a target object into the correct trash can. If it failed to recognize/grasp the object, navigate to a destination, or throw away the object, a negative point was given as penalty to the total score.

As shown in Fig. 3 and Table 2, there was a wide distribution in the points acquired by the robot. If the evaluation had been a one-shot trial, team 2 might have won the competition. Our platform enabled the robot to take part in multiple sessions in a variety of conditions such with different types of objects, environmental lightning conditions, and furniture layouts. Human gestures and behavior in particular should be given to the robots as uniform and fair data. This result shows the advantage and feasibility of our platform from the viewpoint of a statistical evaluation of real human-robot interaction in the RoboCup competition.

4 Task II: Human Navigation

4.1 Task

This test was done to evaluate the ability of a robot to generate natural and friendly language expressions to explain how to get to a destination and where a target object is. The robot has to generate natural language expression in accordance with the location and gaze direction of the human avatar and layout of the environment. In this test, communication is done only through text messages and voice utterances. The robot cannot use any gestures or visual information.

A virtual environment for this task is shown in Fig. 4. The environment consists of four rooms, a kitchen, a bed room, a living room, and a lobby. The robot itself does not exist in the environment because communication is carried out using only text messages and voice utterances.

The procedure of the task is as follows.

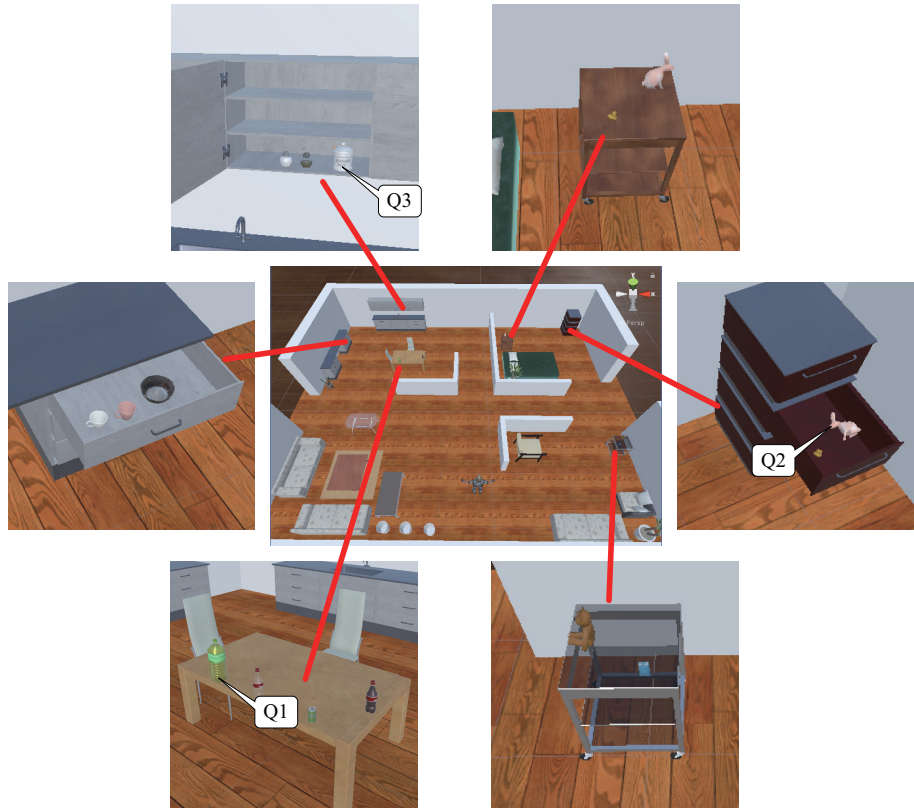


Fig. 4. Environment used in the Human Navigation task.

- The robot system receives the target object information (object ID, location)
- The robot can observe the status (position, posture, gaze direction, etc.) of a human avatar by using an exclusive API in the virtual environment.
- The robot generates appropriate sentences to explain where the target object is. The sentences can be displayed in the HMD or synthesized as voice sound.
- The participant reads or listens to the explanation and controls the human avatar. The robot cannot physically guide the participant such as by pointing.
- The time required to find the target object is measured. The time limitation for each session is 5 minutes. If a participant finds the target or gives up, the robot system proceeds to the next session.
- The session is repeated for several times per participant.

We asked 10 participants to take part in this task by using the immersive virtual reality system. The participants wore an HMD (Oculus Rift CV1) to log in to the human avatar. They controlled the avatar by using a handheld device (Oculus Touch). Three directions given by the robot, Q1, Q2, and Q3,

are shown in Fig. 4. In this primitive investigation, we used the pre-defined utterances shown in Table 3.

Table 3. Instructions used in Human Navigation task

Q1	Please take a green PET bottle on the desk at the kitchen.
Q2	There is a brown chest in the bedroom. Please take a toy bunny from the second drawer from the top.
Q3	There is a cabinet above the kitchen sink. Please take a white sugar pot from the second cabinet from the right.

4.2 Evaluation result

Table 4 shows the result of the experiments for the 10 participants. Figure 5 shows a timeline of a certain session. As shown in the table, almost all of the sessions were completed within one minute. Even though the VR environment was hidden from the participants before the sessions, no one hesitated to go to the right location and successfully grasp the correct target object.

This means that we can manage a competition in which a general audience can take part. Such a competition would provide a fair and open evaluation against the problem for current RoboCup@Home tasks, which depend on just one-shot and non-uniform data. Additionally, another advantage is that the reactions and behavior of avatars can be easily recorded in the system. Quantitative evaluation after a real session can be performed. Objective observation of human behavior should be key in designing the social and cognitive tasks.

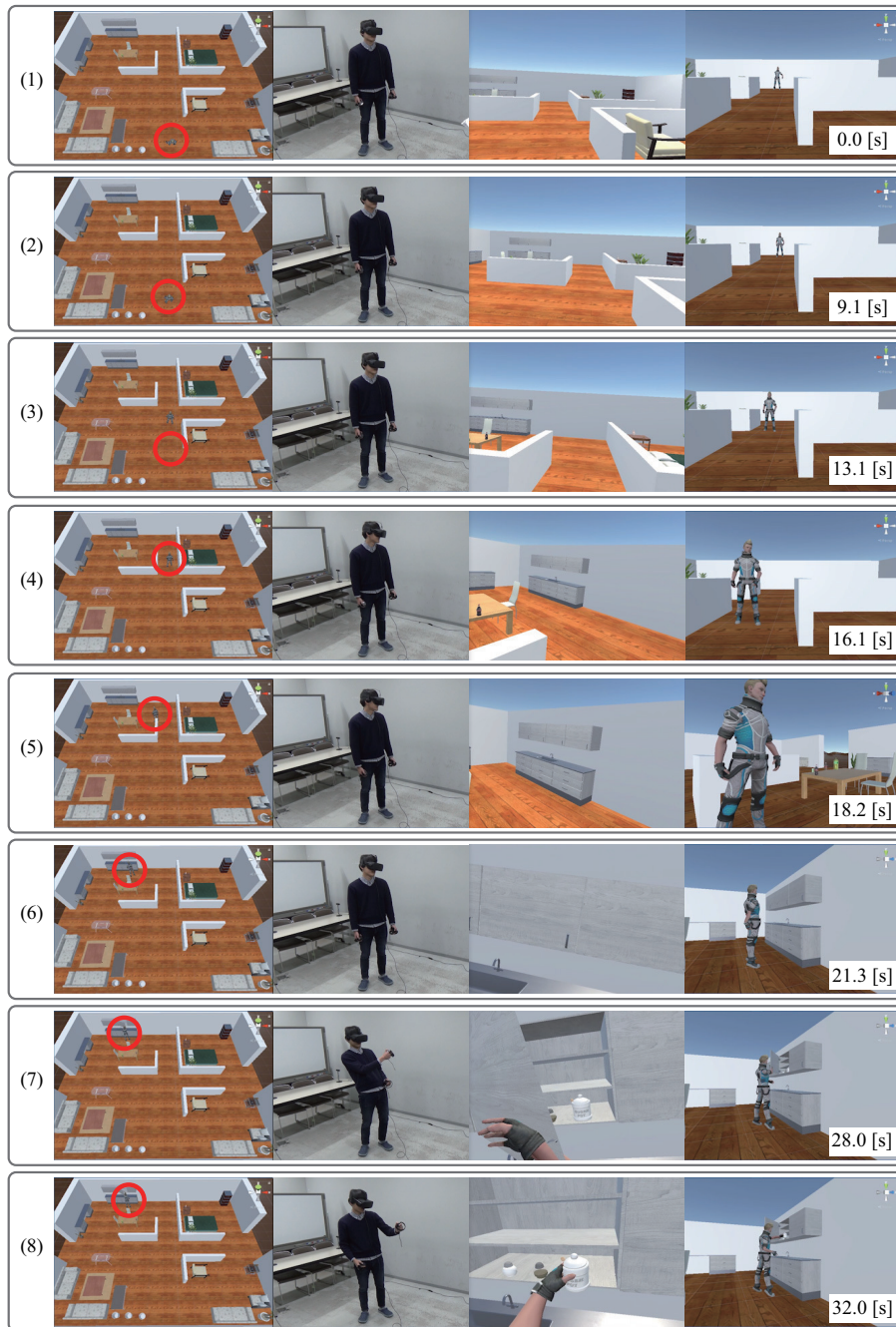


Fig. 5. Timeline of session of Human Navigation task. Four columns in each picture are, from left, whole environment and position of avatar, participant operating avatar, first-person view from avatar, avatar's behavior.

Table 4. Time required to complete task. (unit: [sec]) hint indicates time when subject used replay function.

	Q1	hint	Q2	hint	Q3	hint
A	49		41		42	
B	17		44		32	
C	33		34		30	
D	19		35		34	
E	17	9	26	13	20	
F	27		48		34	
G	37		69		36	
H	35		41		37	
I	19		32		28	
J	40		48		36	
ave	29.3		41.8		32.9	
max	49		69		42	
min	17		26		20	
std	10.6		11.3		5.7	

5 Discussion and Conclusion

We proposed a novel software platform and task design to evaluate high-level cognitive functions in human-robot interaction. The conventional RoboCup@Home tasks tend to focus on physical functions in the real world; however, user behavior and reactions are also an important target of evaluation. Such evaluation is available with the help of the proposed platform.

We carried out two tasks with the proposed platform. The results show that VR simulation can be used to evaluate the quality of human-robot interaction that includes social, physical, and cognitive interaction. Statistical evaluation is difficult in conventional physical human-robot interaction due to limitations in workable time, hardware durability, and provision of fair interaction data. The proposed platform can solve these problems by recording human-robot interaction and crowd-sourcing through immersive VR devices.

Interactive CleanUp and Human Navigation were done at RoboCup JapanOpen 2016 and 2017, respectively. Not only new tasks for cognitive HRI but also conventional tasks such as EGPSR using the proposed platform have been carried out at JapanOpen. This means that a variety of tasks required for RoboCup@Home are feasible on the proposed platform.

In the near future, we believe that the proposed software platform and task design should be carried out for the RoboCup@Home simulation league as well as other simulation leagues.

This system is open source software. Everyone can download the source code from GitHub repository⁴ and get documentation from the SIGVerse wiki⁵.

⁴ <https://github.com/SIGVerse>

⁵ <http://www.sigverse.org/wiki/en/>

References

1. Inamura, T., Tan, J. T. C., Sugiura, K., Nagai, T. and Okada, H.: “Development of RoboCup@Home Simulation towards Long-term Large Scale HRI,” Lecture Notes in Computer Science: RoboCup 2013: Robot World Cup XVII, vol.8371, pp.672–680 (2013)
2. Inamura, T. *et al.*: “Simulator platform that enables social interaction simulation - SIGVerse: SocioIntelliGenesis simulator,” in IEEE/SICE Int. Symp. on System Integration, pp. 212-217 (2010)
3. Codd-Downey, R., Forooshani, P. M., Speers, A., Wang, H., and Jenkin, M.: “From ROS to unity: Leveraging robot and virtual environment middleware for immersive teleoperation,” in IEEE Int. Conf. on Information and Automation, pp. 932–936 (2014)
4. Hu, Y., and Meng, W.: “ROSEUnitySim: development of a low-cost experimental quadcopter testbed using an arduino controller and software,” J. Simulation, vol. 92, Issue. 10, pp. 931–944 (2016)
5. Koenig, N., and Howard, A.: “Design and use paradigms for gazebo, an open-source multi-robot simulator,” in Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2149–2154 (2004)
6. Lewis, M., Wang, J., and Hughes, S.: “USARSim: Simulation for the Study of Human-Robot Interaction,” Journal of Cognitive Engineering and Decision Making, vol. 1, no. 1, pp. 98–120 (2007)
7. Nakaoka, S.: “Choreonoid: Extensible virtual robot environment built on an integrated GUI framework,” in Proc. of IEEE/SICE International Symposium on System Integration, pp. 79–85 (2012)
8. Kanehiro, F., Hirukawa, H., and Kajita, S.: “OpenHRP: Open architecture humanoid robotics platform,” The International Journal of Robotics Research, vol. 23, No. 2, pp. 155–165 (2004)
9. Rohmer, E., Singh, S. P. N., and Freese, M.: “V-REP: A versatile and scalable robot simulation framework,” in IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1321–1326 (2013)
10. Michel, O.: “Webots TM : Professional Mobile Robot Simulation,” International Journal of Advanced Robotic Systems, vol. 1, no. 1, pp. 39–42 (2004)