Long-term Large Scale Human-Robot Interaction Platform through Immersive VR System
–Development of RoboCup @Home Simulator–

*Tetsunari INAMURA and Jeffrey Too Chuan TAN

***Abstract*— 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. In this paper we explain configuration of our simulator platform and feasibility of the system in RoboCup @Home.**

I. INTRODUCTION

One of the important key issues on human-robot interaction (HRI) in real world is elucidation of mechanism of social and physical interaction and embodied design of robots based on its understanding. Aiming such goal, several trials have been proposed such as observation/analysis of human's interaction behavior between humanoid/mobile/pet robots through several different conditions on social/physical interaction[1][2][3]. Hard problems on such HRI experiments are limited opportunity for setting interaction field for the experiments and closed environment in laboratory in universities/institutes. Another problem is huge cost of people and time for large scaled experiments, especially the aim of HRI is learning from demonstration or instruction such as[4].

To cope with the problem, we have proposed a simulator system SIGVerse[5] that enables easy development of HRI experiments in which human beings and virtual robot agents can make social and physical interaction. This system enables arbitrary users to join virtual HRI experiment through the Internet with log-on to the central virtual world. Speech conversation and simple interaction via GUI were basic interaction methods; however, more natural and intuitive interfaces are required to aim embodied physical HRI system. In this paper, we therefore propose immersive interface devices for the SIGVerse simulator to enable intuitive and embodied interaction, and a framework of multiuser connection to share the central virtual world to simulate social interaction field. We also discuss a feasibility of the system to apply to the RoboCup @Home application that is one of the most realistic and large scaled interaction field for HRI.

National Institute of Informatics (NII) / SOKENDAI, Japan

II. BIG DATA AND HRI RESEARCH

We think one of the most important next research issue on HRI is construction of hypothetical interaction model and its evaluation based on big data that is large scaled experience between humans and robots. Typical conventional HRI research is performed in a closed laboratory; however, large scaled interaction experiment similar to pilot program that requires social and embodied interaction enables to provide rich big data to approach to construction of general interaction model of human beings.

Currently, such research platform that can collect big data for HRI does not exist. However, similar concept is proposed in entertainment field. For example, Internet game named akinator[1] can determine which character the player is thinking of by asking him/her a series of simple closed questions. NERO(Neuro Evolving Robotic Operatives) project proposes a research platform[6] in which agents can play a simulation game, store the game play as database, use the playing database machine learning and so on, to promote a research on machine learning. These applications use social interaction and huge playing database; however, the scope of interaction is just a quiz or game. The interaction medium is also limited to just text message. If we aim to realize embodied interaction, these framework is not satisfied.

On the contrary, there is an appropriate application from social and embodied interaction, that is RoboCup @Home[2]. This is robot competition in which the robot agents are evaluated for performance of interaction ability and task achievement ability in daily life situation such as living and kitchen environments. The robot agents have to hold advanced social and embodied interaction functions such as speech recognition, dialogue management, grasping object, navigation, intention understanding via gestures/eye direction, inference of users' instruction from vague expression and so on. Additionally, machine learning technique is also evaluated to adapt unknown environment/situation based on past experiences. This competition is attractive from a view point of high level HRI such as learning user models, perform natural and intuitive interaction and so on; however the cost to develop real robot agents and to execute experiment in real world is too expensive as shown in the section 1.

Here, our proposal is to apply our simulator platform named SIGVerse to this RoboCup @Home competition.

[1] http://en.akinator.com/

[2] http://www.ai.rug.nl/robocupathome/

From the next section, we explain the detail strategy of the application.

III. THE SIGVERSE: SOCIOINTELLIGENESIS SIMULATOR

In this section, brief overview of the SIGVerse system[5] is shown.

A. Dynamics simulation

The Open Dynamics Engine (ODE)[3] is used for dynamic simulations of interactions between agents and objects. Basically, the motions of each agent and object are calculated by the dynamics engine, but the user can control the calculations to reduce simulation costs. A switch flag can be set for each object and agent to turn off the dynamics calculations if required.

B. Perception simulation

This system can provide the senses of vision, sound, force, and touch. OpenGL is used for visual simulations, to provide each agent with a pixel map that is a visual image derived from the viewpoint and field of view of that agent. In this case, the perception simulation has several levels that control the abstract level of perception.

At a highly abstract perception level, the user is sent symbolized visual information, which comprises data such as the color, shape, size, and position of each object within sight, together with characteristic information on the name and ID number of each object. The visual perception processing also considers occlusions so that if an object is behind another object, the perception processing omits the hidden object from the list.

Regarding the sound simulation, raw sound patterns are transferred among agents. Simulation of acoustic field and reflection of sound are not implemented due to computation cost. However, the power of the sounds can be reduced according to the distance between sound sources. Another function has been realized that agents who exist in a certain distance from the sound source can hear the sound. Not only raw sound data, but also text messages which are converted from voice via speech recognition systems can be transferred.

For the sense of touch, it is possible to acquire force and torque information between objects that has been calculated mainly by ODE.

C. Communication

In simulating the sense of hearing, every agent can make communications with audio data. I also enabled the effect of the volume of sound attenuating in inverse relationship with the square of distance, based on a setting of a condition that the voice emitted by an agent becomes more difficult to hear with distance. It is also possible to set the system so that only voices within a certain threshold distance are acquired.

With this system, not only can agents within the virtual environment communicate with each other, it is also possible to provide a function that enables interactions between the virtual environment and users in the real world.

[3]http://www.ode.org

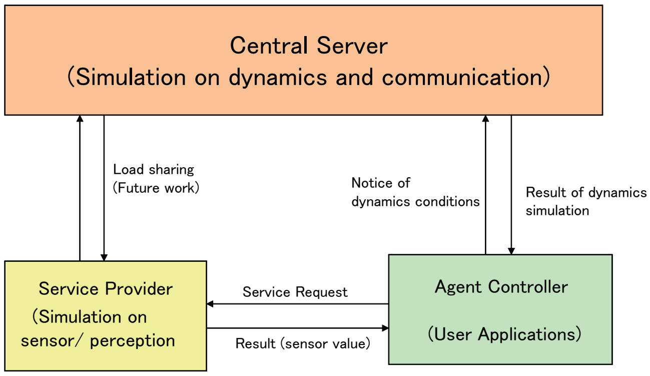

Fig. 1. Software configuration of SIGVerse

As related works, there are several projects[7][8] in which human agents act in virtual social world based on Second Life-wise framework. However, dynamics and perception simulation is hard to implement easily.

D. Configuration of simulator software

This simulator consists of server and client systems. Dynamics calculations are mainly performed on a central server system. Bodies that use perception and perform actions are called ”agents”, for example robots and human avatars. Autonomous actions of the agents should be designed by C++ APIs shown in Table 1. In near future, ROS[4] and OpenRTM[5] will be connected as a agent's behavior programming method. The avatars' behaviors are programmable using the APIs; they can also act on the basis of instructions given to them by operators in real time.

The configuration of the SIGVerse software is shown in Fig.1.

IV. FUNCTIONS REQUIRED IN ROBOCUP @HOME AND NEW FUNCTIONS OF THE SIGVERSE

In this section, we introduce required functions in the RoboCup @Home, and propose enhanced interface devices to enable virtual RoboCup @Home competition in the SIGVerse system.

A. Frequent behavior in RoboCup @Home tasks

Tasks in the RoboCup @Home competition is defined in the rule-book[9]; the name of tasks are *Follow Me, Clean Up, Enhanced Who is who* and so on. In the *Follow Me*, robots have to recognize face image of users, track the face image and follow the walking users. In the *Clean Up*, robots have to grasp target trashes instructed by users, and throw away the trashes. In the *Enhanced Who is Who*, robots have to bring a drink cups/bottles to unknown users, and hand over the drinks. To realize these tasks, development of many basic functions are required such as receiving users' instructions via speech and gestures, understanding the meaning of the instruction by speech recognition, image processing for captured picture by cameras, grasping of object by end effectors,

[4]http://www.ros.org/wiki/
[5]http://www.openrtm.org/

Fig. 2. A user joining a virtual world through HMD. (Used environment is the same as Fig.4)

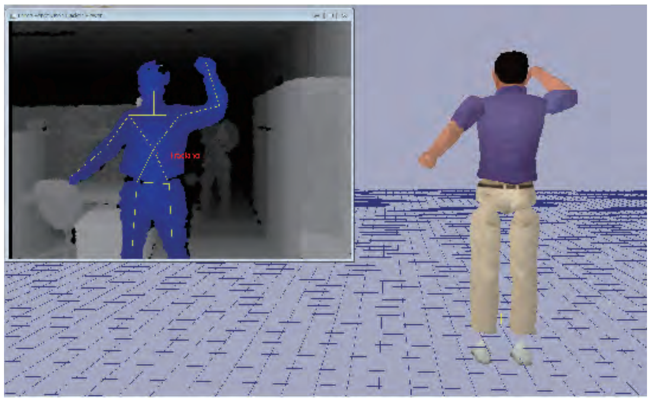

Fig. 3. Whole body control of an avatar by KINECT sensor

navigation by wheels, dialogue management system and so on. Basically, these basic functions can be simulated in the SIGVerse world.

B. Introduction of immersive interface for RoboCup @Home

However, several limitations existed on a problem how to project the user's motion and behavior into an avatar in the virtual world. Conventional SIGVerse interface used a large motion capturing room which has an omni-directional surrounding display[10] to project the user's motion and to show the virtual world to the user. However, such inflexible large equipment is not suitable for a competition such as RoboCup @Home in which many arbitrary users join, because many users cannot share such massive and expensive equipment. We here propose a cheap, flexible and immersive interface for the SIGVerse to enable social and embodied interaction with virtual agents.

1) Projection of SIGVerse world to HMD: In the proposed system, a head mounted display Z800 3DVisor (eMagin Production) was used. This HMD has motion detector for pan and tilt direction. Detected motion is transferred to SIGVerse system to control an avatar. Of course the HMD can display sequence of images that are captured by avatar's eye according to the head motion. Fig.2 shows an overview of the real user and projected scene image on the HMD.

2) Control of whole body motion of avatars by Kinect: Microsoft KINECT can be connected to the latest SIGVerse client terminal. As well as the motion detector on a HMD, motion pattern measured by the KINECT is transferred to the server to control whole body motion of the avatar. This function enables a virtual behavior which uses both of speech command such as "Please take the drink" and pointing gestures. Not only pointing gesture, but the avatar can also grasp virtual objects with showing grasping motion in front of KINECT. Motions of fingers are difficult to measure by KINECT, the system thus provide grasp/release command to control grasping status of the avatar. The grasp/release commands are often given by speech instruction by users. Fig.3 shows a screen-shot of avatar controller via KINECT.

Those immersive interfaces can be connected to the SIGVerse client system as plug-in modules. According to the progress of RoboCup @Home simulation league, additional proposals of immersive interfaces from participants are expected. Therefore, the SIGVerse client has a plug-in system to accept arbitrary interface systems, such as users' original speech recognition systems, sound source detection, face recognition systems, haptic interfaces, and so on.

C. Limitations of the RoboCup @Home simulator interface

In the *Follow Me* task, robots have to track a user's face image and follow the walking user. But the SIGVerse client system is just a computer, not a mobile robot typically used in the RoboCup @Home. Therefore, one of the limitations is that the users have to behave in small observation area which is covered by fixed sensors such as KINECT, camera, distance sensor, and so on. The SIGVerse can use motion capture systems such as [10] to reflect the humans' walking around behavior; however, such motion capture systems are expensive and not easy to be used by a lot of participants of the RoboCup @Home.

In the *Clean UP* and *Enhanced Who is who* tasks, robots have to grasp objects such as drinks and bring/hand over it to destination; however, the SIGVerse system doesn't provide accurate simulation of physical grasping behavior with considering friction on soft and flexible material. This problem is still hard issue even in manipulation research. Therefore, the proposed system skips the manipulation simulation currently; the system manages binary status such as grasping/not grasping. As well as the object manipulation, hand over from robot agents to human avatars is only performed as changes of status. The user cannot feel the reaction force.

The SIGVerse system however can use haptic interface systems such as Phantom Desktop[5]. Thus, haptic interface is one of the available extension for RoboCup @Home simulation. As well as the motion capture systems, the haptic interface might be adopted as an interface if the cost is reduced.

We do not aim to realize all the existing tasks of RoboCup @Home in the SIGVerse simulator; but aim to propose advanced interaction scenarios that require high level agent's function such as decision making, machine learning from conversation, inference of future event, and so on. We propose an example of such scenarios in the next section.

V. Applications to RoboCup @Home tasks

A. Cooking with Me task

Not only tasks described on the rule-book, but also several tasks are feasible on SIGVerse system and appropriate for the scope of RoboCup @Home. One of the candidate tasks is cooperative cooking task. The cooperative cooking task is one of the major HRI task; several related works have been proposed in robotics field[11][12] and human-agent interaction field[13]. This task domain requires recognition of human's behavior, realtime planning according to receipt, dialogue management and so on, those are also required and fit for RoboCup @Home. So far, researcher have to tackle with real-world oriented problems such as sensor technology to cope with liquid or soft materials, manipulation of complex and soft objects and so on, even though the researcher's interest is aiming high level interaction function.

Here, we propose setting the cooperative cooking task as a representative task of simulation league of RoboCup @Home. Another reason to choose the cooking task is that the frequency of walking around a field is low in the cooking task and expected area of human's motion is often limited within a few steps. Since typical human's behaviors are cutting foodstuffs, grasping dishes/cups, operating microwaves and so on, observation of such behaviors can be focused on upper-body motion and it is easily accomplished by KINECT. Figure 4 shows an example field in which a robot agent and an avatar are cooperating to make an Okonomiyaki, which is a kind of Japanese pancake.

Task representation of the cooking is described as state transition diagram. Each node is connected to another node with an arc which is corresponded to an action. Each action has conditions to be activated, and effect to the virtual world as output. Those actions are designed previously and given to the robot and users. The task for participants to the competition is to design the state transition diagram manually or to develop machine learning technique to make the robot acquire the state transition diagram through human-robot interaction. Another task is to develop effective human-robot interaction strategies. For example, if the user doesn't take action for the cooking, the robot may ask the user to do something in parallel. The robot also should search and find suitable action when the user already start an action. Such action selection algorithm is another target to make the system more friendly. An original state transition diagram is given as CSV file format. The participants are asked to write their own program to refer/modify the state transition diagram shown in Figure 4.

Final task is understanding of the users' reference object using natural language and gesture instruction. Gesture and speech are basic interface for the users. If users want to ask the robot to move/handle/take objects, users have to speak such as "Please bring the dish to the dining table" with pointing the dish. The robot should recognize which dish is the target using the posture of the pointing gesture and surrounded situation; however vague pointing and speech are expected. In such case, additional interaction would be required to clear the uncertainty using confirmation/question.

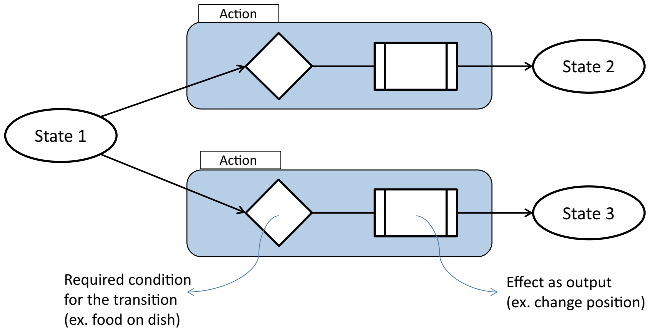

Fig. 4. A state transition diagram for the cooking task. Each action consist of required conditions (diamond block) and output effects (pre-defined procedure as rectangle).

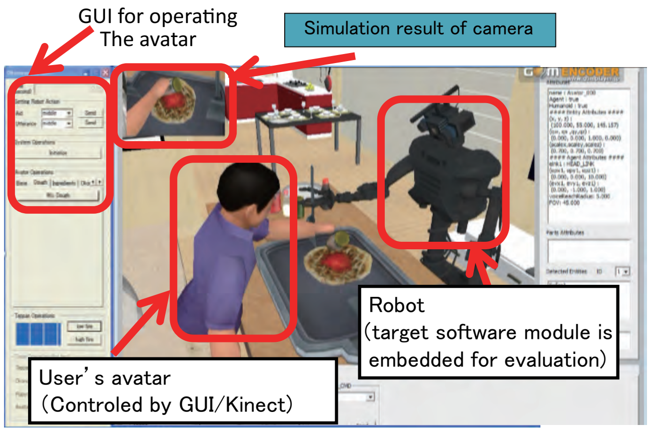

Fig. 5. Screen-shot of the Cooking with Me task

This dialogue management is one of the high-level interaction functions which is fit for the simulation based HRI.

Basically, each action such as take, move, bring, flip, drop and so on are pre-defined including grasping/motion planner which can generate suitable motion trajectory as long as the target object is given.

We performed primary experiments on two cases: one in which the user performed all of the steps through the GUI, and one in which the robot did suitable parts of the operator's work instead. In the first case in which the operator did all of the work, the task required 194[sec] before it was finished, but in the second case involving cooperation, the task took one minute 58[sec] to complete. The required time in this experiment would be one of the evaluation criterion of RoboCup @Home competition.

B. Environment of RoboCup @home

As shown in the above examples, RoboCup @Home often uses living and kitchen environment. The latest SIGVerse system provides open and public components to build living and kitchen environment. Users can create their favorite environment with a design application. Figure 5 shows the overview of open components.

Fig. 6. Objects in kitchen (above) and living (below) environment as open and public components

Fig. 7. SIGVerse Web: the development environment on web browser

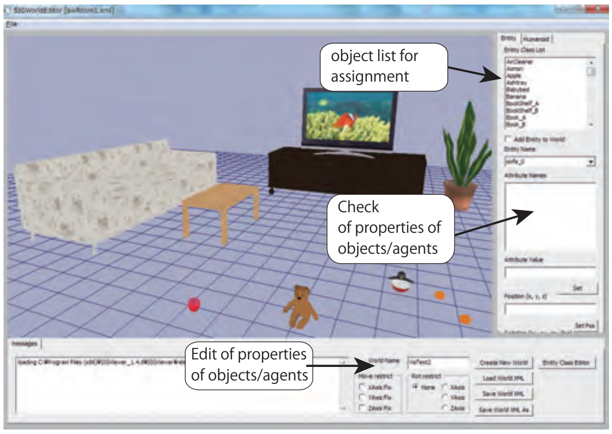

Fig. 8. Design of virtual world using the SIGWorldEditor

VI. Development Environment of SIGVerse

This section describes usage of development environment of SIGVerse.

A. Setup of SIGVerse

SIGVerse is a client/server system that consists of a Linux server and a Windows client application. Information of all the agents if controlled by the Linux server. Each user connects to the server system through the Windows client system. Everyone can download and install the client application from download site[6].

B. Development environment for users

In the SIGVerse system, programs which determine behavior of agents are called as *controller*. The controller should be compiled as dynamic link library on the Linux SIGVerse server. The controller is attached to SIGVerse server process according to a configuration file which is written by users. Thus, every user shall setup the Linux server for the SIGVerse; compile the controller on the server. However, for general users who are interested in the RoboCup @Home competition might not be interested in server management. For such users, our institute (National Institute of Informatics) opens a development environment system through web interface, which name is SIGVerse Web. Figure 6 shows a screen-shot of the SIGVerse Web. Operation for the development such as edit and compilation of source codes can be realized by web based interface. Since classical terminal login to the SIGVerse server is not required; beginner of Linux system can easily develop the controller through only GUI on web browser.

Source codes for sample application are also available at SIGVerse tutorial[7]. The tutorial has more than 15 sample programs such as 1) Controller of mobile robots, humanoid robots and human avatar; 2) sensor emulator such as distance sensor; and 3) usage of Kinect and head mount display for immersive VR system.

A configuration file is required to determine the virtual world for each application, such as position of agents and objects, property of objects and so on. The configuration should be described by XML file. Users can also design the virtual world using simple and easy interface application named `SIGWorldEditor`. This application is also available from the download site. Figure 7 shows a screen-shot of the world design application.

C. Execution of simulation

After the configuration and design of the virtual world, users can execute the simulation through client software named `SIGViewer`. The time in the virtual world can pass

[6]http://www.sigverse.org/

[7]http://www.sociointelligenesis.org/SIGVerse/index.php?Tutorial

as the time in our real world when the users connect to the virtual avatar via interface devices. Of course, the time in the virtual world can pass quickly compared with the time in the real world.

SIGVerse has text based chat and speech recognition/synthesis as standard user interface. Users can communicate with virtual avatar and robots via those interface. Additionally, Kinect, joystick controller motion capture system are able to be connected to control the agents.

VII. Discussion and Conclusion

In this paper, we focused on social/embodied interaction and adaptation/learning method based on big data as important research issues on next step of Human-Robot Interaction. We also showed that implementation of simulator platform in which everyone can easily join social interaction experiments will be a breakthrough to promote Human-Robot Interaction research.

Development of an immersive interface for the SIGVerse system is introduced. We also showed that an effective competition task can be provided for RoboCup @Home with multi-user connection environment on the Internet. The cooperative cooking task for RoboCup @Home is just a one of the milestones. The interaction is currently one-to-one; however it is easy to develop many-to-many human-agent interaction scenario based on the multi-user connection system. We are also planning to apply the proposed system to sport coaching system[14] as one of the HRI system.

However, the proposed system still has several limitations. The user cannot feel presence and texture of robot agents. Reaction as physical interaction is also difficult to be shown to the users. Embodied interaction realized by the proposed system is limited to kinematic whole body motion and movement of eye direction. Thus, the coverage of the proposed system is simulating of interactions that use kinematic motions/gestures, eye direction, speech. We think tasks on RoboCup @Home have moderate level of complexity even though the embodied interaction was limited.

Currently, RoboCup @Home does not have simulation league. We are preparing to hold RoboCup @Home simulation competition at RoboCup Japan Open held in 2013 at the earliest. Latest information is announced at the web-site of the SIGVerse (http://www.sigverse.org/).

Acknowledgment

This research is supported by Grand challenge program of NII, and JSPS Kakenhi grant No.2300077, 24700199.

References

[1] Vijay Chidambaram, Yueh-Hsuan Chiang, and Bilge Mutlu. Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In *Proc. of the 7th ACM/IEEE Int'l Conf. on Human-Robot Interaction*, pp. 293–300, 2012.

[2] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. Personalization in hri: A longitudinal field experiment. In *Proc. of the 7th ACM/IEEE Int'l Conf. on Human-Robot Interaction*, pp. 319–326, 2012.

[3] Anja Austermann and Seiji Yamada. Teaching a pet-robot to understand user feedback through interactive virtual training tasks. *Autonomous Agents and Multi-Agent Systems*, Vol. 20, No. 1, pp. 85–104, 2010.

[4] Komei Sugiura, Naoto Iwahashi, Hisashi Kawai, and Satoshi Nakamura. Situated spoken dialogue with robots using active learning. *Advanced Robotics*, Vol. 25, No. 17, pp. 2207–2232, 2011.

[5] Tetsunari Inamura. Simulator platform that enables social interaction simulation –SIGVerse: SocioIntelliGenesis simulator–. In *IEEE/SICE International Symposium on System Integration*, pp. 212–217, 2010.

[6] Risto Miikkulainen. Creating intelligent agents in games. *The Bridge*, pp. 5–13, 2006.

[7] Anette von Kapri, Sebastian Ullrich, Boris Brandherm, and Helmut Prendinger. Global lab: an interaction, simulation, and experimentation platform based on "second life" and "opensimulator". In *Proc. Pacific-Rim Symp on Image and Video Technology (PSIVT'09)*, 2009.

[8] Magnus Johansson and Harko Verhagen. Massively multiple online role playing games as normative multiagent systems. In *Normative Multi-Agent Systems*, No. 09121 in Dagstuhl Seminar Proceedings, 2009.

[9] Robocup @home rules & regulations, 2012. Version 2012, Revision 286:288, http://www.ais.uni-bonn.de/∼holz/2012_rulebook.pdf.

[10] Ohhoon Kwon and Tetsunari Inamura. Surrounding display and gesture based robot interaction space to enhance user perception for teleoperated robots. In *Int'l Conf. on Advanced Mechatronics*, pp. 277–282, 2010.

[11] Yuta Sugiura, Daisuke Sakamoto, Anusha Withana, Masahiko Inami, and Takeo Igarashi. Cooking with robots: Designing a household system working in open environments. In *Proc. of CHI2010*, pp. 2427–2430. ACM, 2010.

[12] Mario Bollini, Stefanie Tellex, Tyler Thompson, Nicholas Roy, and Daniela Rus. Interpreting and executing recipes with a cooking robot. In *Proc. of International Symposium on Experimental Robotics (ISER)*, 2012.

[13] Sven Reichel, Timo Muller, Oliver Stamm, Fabian Groh, Bjorn Wiedersheim, and Michael Weber. Mampf: An intelligent cooking agent for zoneless stoves. In *7th International Conference on Intelligent Environments*, pp. 171–178, 2011.

[14] Keisuke Okuno and Tetsunari Inamura. Motion coaching with emphatic motions and adverbial expressions for human beings by robotic system. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3381–3386, 2011.