



Experimental Observation of Nodding Motion in Remote Communication Using ARM-COMS

Teruaki Ito¹(✉), Hiroki Kimachi², and Tomio Watanabe³

¹ Graduate School of Technology, Industrial and Social Sciences,
Tokushima University, 2-1 Minami-Josanjima, Tokushima 770-8506, Japan
tito@tokushima-u.ac.jp

² Graduate School of Advanced Technology and Science,
Tokushima University, 2-1 Minami-Josanjima, Tokushima 770-8506, Japan
c501732024@tokushima-u.ac.jp

³ Faculty of Computer Science and System Engineering,
Okayama Prefectural University, 111 Tsuboki, Souja, Okayama 719-1197, Japan
watanabe@cse.oka-pu.ac.jp

Abstract. Considering the critical issues of remote communication, this study proposes an idea of remote individuals' virtual connection through augmented tele-presence systems called ARM-COMS (ARm-supported eM-bodied COMMunication Monitor System). Several ideas of robot-based remote communication systems have been proposed to challenge the telepresence issue of remote participants. However, it does not cover the issue of relationship. An idea of robotic arm-typed system and/or an idea of anthropomorphization draw researchers' attentions to challenge the lack of relationship with remote participants. However, usage of the human body movement of a remote person as a non-verbal message, or cyber-physical media in remote communication is still an open issue. Under these circumstances, this paper describes the system configuration of ARM-COMS based on the proposed idea and discusses the feasibility of the idea using the experimental observations.

Keywords: Cyber-physical communication media · Embodied communication
Augmented tele-presence robotic arm manipulation · Face detection

1 Introduction

TV phone was regarded as a dream for communication tools in SF movies in the old days. However, a smartphone-based video communication tool is now one of the convenient popular tools freely available to mostly everybody [1]. Supporting by ICT (Information and Communication Technology) technologies, further enhancement of better communication is expected. In the meantime, this tool addresses the two types of critical issues, which are the lack of tele-presence feeling and the lack of relationship feeling in remote video communication [5] as opposed to a face-to-face communication.

Several ideas of robot-based remote communication systems have been proposed as one of the solutions to the former issue; these robots include physical telepresence robots [9, 21, 22]. Anthropomorphization [14] is another new idea to show the telepresence of a remote person in communication system. Remote communication can

be basically supported by the primitive functions of physical tele-presence robots, such as a face image display of the operator [15], as well as tele-operation function such as remote-drivability to move around [10], or tele-manipulation [10]. However, there are still an open issue to be studied to narrow the gap between robot-based video communication and face-to-face one.

The second issue in the lack of relationship-type feeling in remote video communication is another big challenge. Recently, an idea of robotic arm-type systems draws researchers' attention [25]. For example, Kubi [13], which is a non-mobile arm type robot, allows the remote user to "look around" during video communication by way of commanding Kubi where to aim the tablet with an intuitive remote control over the net. Furthermore, an idea of enhanced motion display has also been reported [16] to show its feasibility over the conventional display. However, the usage of the human body movement of a remote person as a non-verbal message is still an open issue.

This study proposes an idea of human-computer interaction through remote individuals' connection with augmented tele-presence systems called ARM-COMS (ARm-supported eMbodied COmmunication Monitor System) [6, 7, 14]. The challenge of this idea is to use the human body movement of a remote person as a non-verbal message for sharing the connected communication, and to implement a cyber-physical media us-ing ACM-COMS for connected remote communication [8].

2 Overview of ARM-COMS (ARm-supported eMbodied COmmunication Monitor System)

2.1 System Overview of ARM-COMS

Considering the physical entrainment motion in human communication [24], this research challenges these two issues mentioned in the Sect. 1 by the idea of ARM-COMS (ARm-supported eMbodied COmmunication Monitor System) [6]. This paper focuses on the nodding motion as a non-verbal message contents in remote communication using ARM-COMS. Figure 1 shows the system overview of ARM-COM for the experiment in this study. Face detection procedure of a prototype of ARM-COMS is based on the algorithm of FaceNet [20], which includes image processing library OpenCV 3.1.0 [17], machine learning library dlib 18.18 [3], and face detection tool OpenFace [18] which were installed on a control PC with Ubuntu 14.04 [23] as shown in Fig. 2. Using the input image data from USB camera, landmark detection is processed.

ARM-COMS is composed of a tablet PC and a desktop robotic arm. The table PC in ARM-COMS is a typical ICT (Information and Communication Technology) device and the desktop robotic arm works as a manipulator of the tablet, of which position and movements are autonomously manipulated based on the behavior of a human user who communicates with remote person through ACM-COMS. This autonomous manipulation of ARM-COMS is controlled by the head movement, which can be recognized by one of the typical portable sensors, such as a magnetic sensor, gyro-sensor, motion capturing sensor, or a typical cameras, such as Kinect [11] sensor, or a general USB camera.

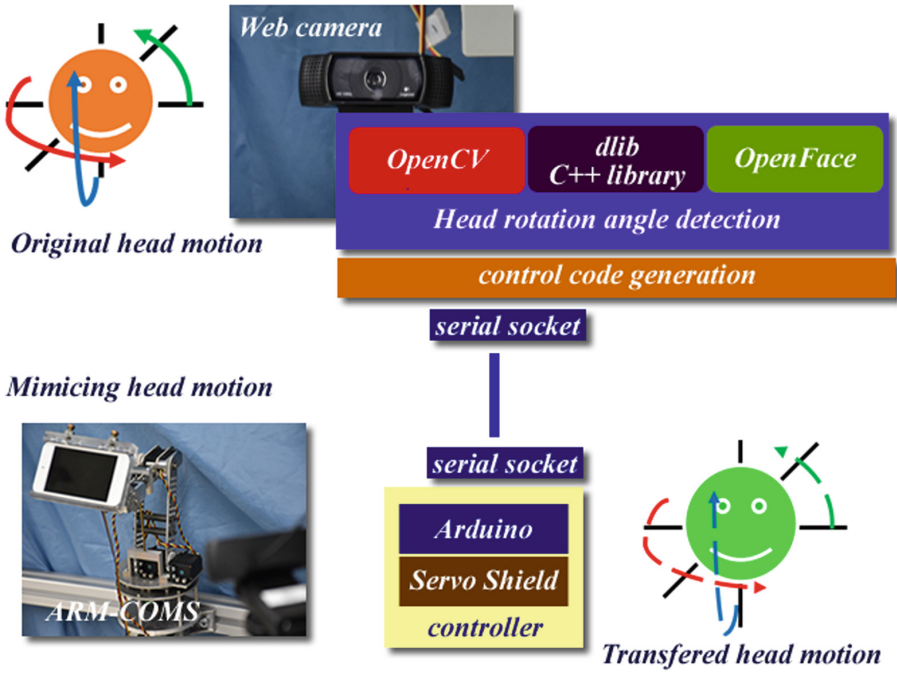


Fig. 1. System architecture of ARM-COMS prototype

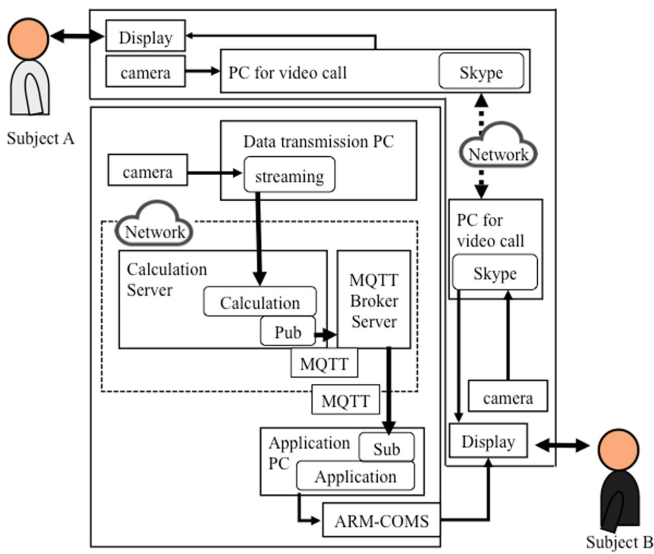


Fig. 2. Network-based configuration of ARM-COMS communication

2.2 System Configuration of ARM-COMS for Network Usage

ARM-COMS is configured to implement network communication as shown in Fig. 2. Head motion of Subject A is used as a non-verbal communication to ARM-COMS which interact with Subject B. Video communication itself was performed by a typical software (Skype). However, the head motion image data is processed by the face detection algorithms mentioned in the Sect. 2.1, which was used to trigger the motion of ARM-COMS installed at the site of subject B.

3 Experimental Comparison

3.1 Experimental Configuration for Nodding Observation

Based on the system configuration shown in Fig. 2, three types of experimental setups were configured, which include (a) face-to-face communication, (b) video communication, and (c) ARM-COMS communication as shown in Fig. 3. The detailed experimental setups are shown in Figs. 4, 5 and 6.

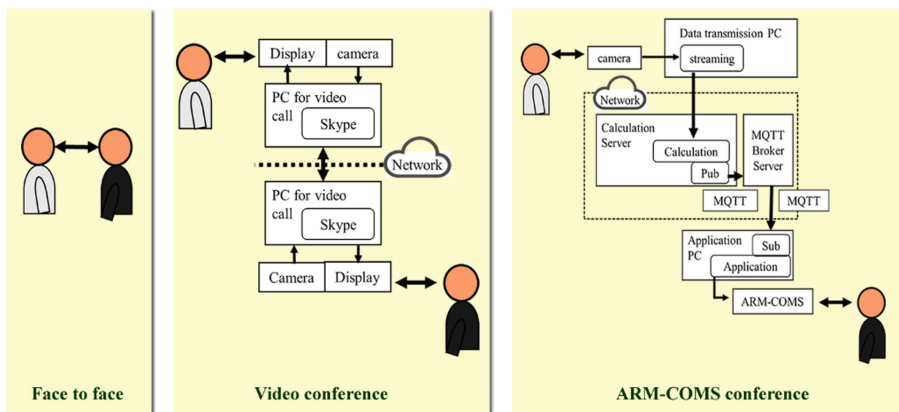


Fig. 3. Three types of experimental setup

Communication experiments were conducted by 8 subjects, which were composed of a pair to make communicate in three types of setups for a short conversation with maximum 2 min conversation in the procedure below.

Experimental procedure:

- Step 1: Subject A and B are positioned to see each other
- Step 2: Subject A and B start nodding in the beginning of conversation.
- Step 3: Subject A and B start short conversation on a topic of breakfast menu.
- Step 4: Subject A and B end conversation by nodding greeting.

3.2 Experiments for Face-to-Face Communication

Figure 4 is experimental setup for face-to-face communication. Head-motion of a human subject is detected and traced according to the short conversation. One magnetic receiver (Fastrak RX-2 [4]) is attached to the head of human subject A and another magnetic receiver B is attached to the other subject.

Figure 5 shows a result of this experiment, which shows the clear correspondence to the nodding interaction between the two subjects.

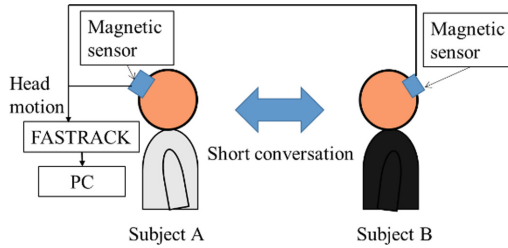


Fig. 4. Experimental configuration of face-to-face communication

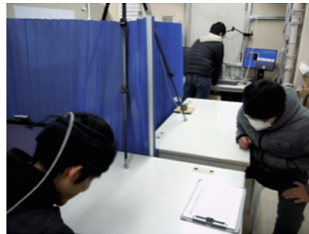


Fig. 5. An experimental result of face-to-face communication

3.3 Experiments for Video Communication

Figure 6 shows the experimental setup for video communication. Head-motion of Subject A and B was detected and traced during the short conversation using magnetic sensor and video imaging as well as gaze point tracking sensor. A general USB camera (Buffalo) captures the image of human subjects during the experiments. A desktop PC

(Windows 7/64) was used for the data collection, whereas a laptop pc (Ubuntu 14.04) was used for ARM-COMS control.

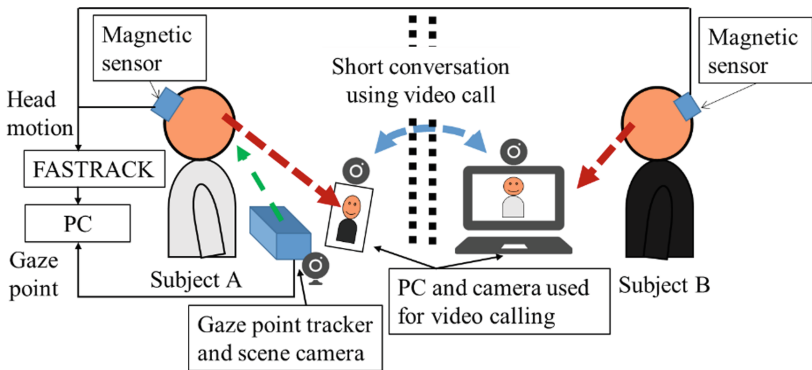


Fig. 6. Experimental setup for video communication

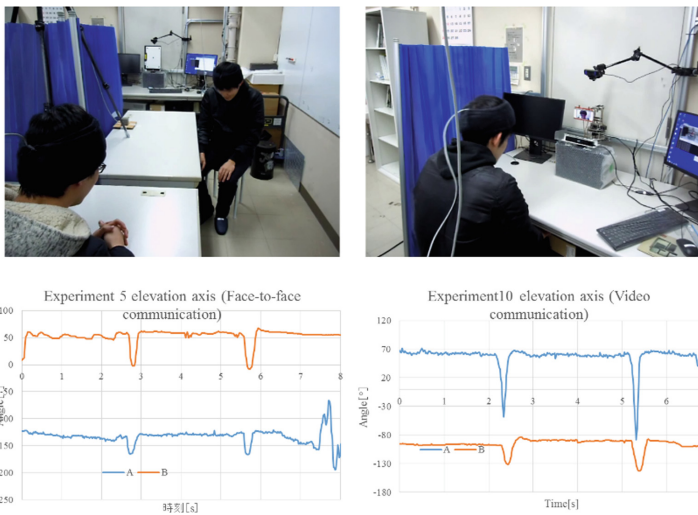


Fig. 7. Experiment 5 & 10

Figure 7 shows some results of this experiment, which shows the clear correspondence to the nodding interaction between the two subjects, both in face-to-face conversation and video conversation.

3.4 Experiments for ARM-COMS Communication

Figure 8 is experimental setup for ARM-COMS communication. Head-motion of Subject A and B was detected and traced during the short conversation using magnetic sensor and video imaging as well as gaze point tracking sensor.

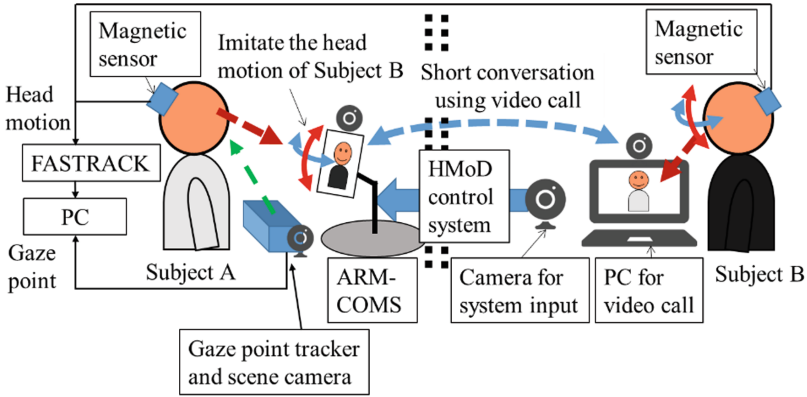


Fig. 8. Experimental setup for ARM-COMS communication

Figure 9 shows some results of this experiment, which shows the clear correspondence to the nodding interaction between the two subjects, both in face-to-face conversation, video communication and ARM-COMS conversation. However, the result did not show any significant difference between video communication and ARM-COMS communication. Further experiments were required.

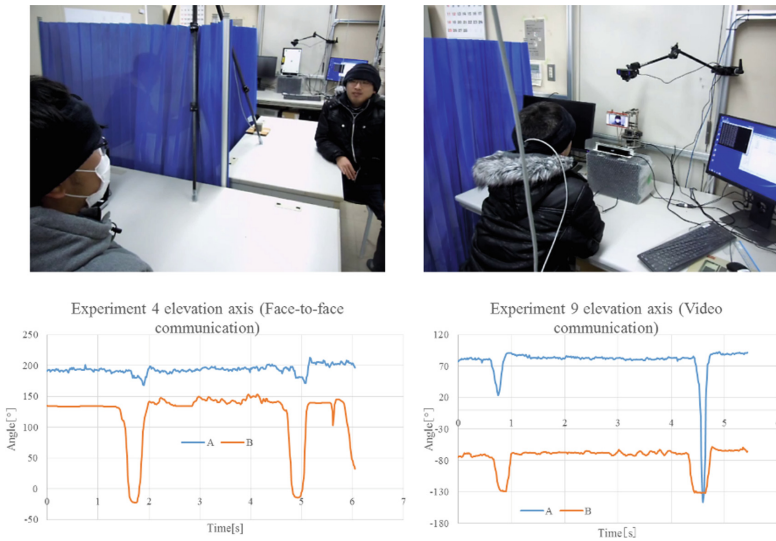


Fig. 9. Experiment 4 & 9

3.5 Results and Discussion

Three types of experiments were conducted to study the feasibility of ARM-COMS communication, namely, face-to-face communication, video communication, and ARM-COMS communication. As shown in the experimental results in the Sects. 3.1, 3.2, 3.3 and 3.4, the experimental setups worked well and three types of experiments were all conducted. Therefore, it could be mentioned that the systems were well configured to implement the idea of this research.

Nodding motion in conversation is very common for Japanese culture, whereas it is not so common for other cultures. Therefore, the subjects including Malaysian and Chinese as well as Japanese attended to the experiments in order to see the difference. The basic instruction was given to the subjects as mentioned in the Sect. 3.1 before the experiment. However, the naturalness of nodding was not similar in subjects, of which difference was clearly recognized by the authors, but was not well analyzed by the experimental data.

According to the head motion data, there was no significant difference between face-to-face communication and video communication for all subjects. Natural nodding style of Japanese subjects was observed both in face-to-face and video communication, which was recognized by head tracking data analysis. Unnatural nodding gesture of non-Japanese subjects was observed both in face-to-face or video communication. However, the unnaturalness of nodding gesture was not recognized by the collected data of the experiments. Since the nodding style is another issue to be studied to show the feasibility of ARM-COMS idea, further design of experimental setup should be considered.

In addition to head motion tracking by magnetic sensor, eye-tracking measurement was also conducted to trace the eye movement [12] during the conversation to see the difference between a typical video communication and ARM-COMS communication. However, eye tracking of subjects could not be traced during nodding motion because the gaze point of subject disappeared out of sight from eye tracker. Therefore, gaze tracking data were not well utilized in the experiments. Further design of experimental setup should be considered.

4 Concluding Remarks

This study proposed an idea of human-computer interaction through remote individuals' connection with augmented tele-presence systems called ARM-COMS (ARm-supported eMbodied COmmunication Monitor System). The challenge of this idea is to use the human body movement of a remote person as a non-verbal message for sharing the connected communication, and to implement a cyber-physical media using ACM-COMS for connected remote communication. Based on the implemented communication platform prototype presented in this paper, three types of experiments were conducted to study the feasibility of the proposed idea. The configuration of the prototype worked well for the experiment. However, Further consideration for design experiments is required to collect measurement data, which will be used for feasibility analysis of the proposed idea.

Acknowledgement. This work was supported by JSPS KAKENHI Grant Numbers JP16K00274. The author would like to acknowledge all members of Collaborative Engineering Labs at Tokushima University, and Center for Technical Support of Tokushima University, for their cooperation to conduct the experiments.

References

1. Abowdm, D.G., Mynatt, D.E.: Charting past, present, and future research in ubiquitous computing. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **7**(1), 29–58 (2000)
2. Bertrand, C., Bourdeau, L.: Research interviews by Skype: a new data collection method. In: Esteves, J. (ed.) *Proceedings from the 9th European Conference on Research Methods*, pp. 70–79. IE Business School, Spain (2010)
3. Dlib C++ library. <http://dlib.net/>
4. FASTRK. <http://polhemus.com/motion-tracking/all-trackers/fastrak>
5. Greenberg, S.: Peepholes: low cost awareness of one's community. In: *1996 Conference Companion on Human Factors in Computing Systems: Common Ground*, Vancouver, British Columbia, Canada, pp. 206–207 (1996)
6. Ito, T., Watanabe, T.: Three key challenges in ARM-COMS for entrainment effect acceleration in remote communication. In: Yamamoto, S. (ed.) *HCI 2014. LNCS*, vol. 8521, pp. 177–186. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07731-4_18
7. Ito, T., Watanabe, T.: ARM-COMS for entrainment effect enhancement in remote communication. In: *Proceedings of the ASME 2015 International Design Engineering Technical Conferences & Computers and Information Engineering Conference (IDETC/CIE2015)*, August, Boston, USA, no. DETC2015-47960 (2015)
8. Ito, T., Watanabe, T.: Motion control algorithm of ARM-COMS for entrainment enhancement. In: Yamamoto, S. (ed.) *HIMI 2016. LNCS*, vol. 9734, pp. 339–346. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40349-6_32
9. Kashiwabara, T., Osawa, H., Shinozawa, K., Imai, M.: TEROOS: a wearable avatar to enhance joint activities. In: *Annual Conference on Human Factors in Computing Systems*, pp. 2001–2004 (2012)
10. Kim, K., Bolton, J., Girouard, A., Cooperstock, J., Vertegaal, R.: TeleHuman: effects of 3D perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In: *Proceedings of CHI2012*, pp. 2531–2540 (2012)
11. Kinect. <https://dev.windows.com/en-us/kinect>
12. Krafska, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., Torralba, A.: Eye tracking for everyone. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016)
13. Kubi. <https://www.revolverobotics.com>
14. Osawa, T., Matsuda, Y., Ohmura, R., Imai, M.: Embodiment of an agent by anthropomorphization of a common object. *Web Intell. Agent Syst.: Int. J.* **10**, 345–358 (2012)
15. Otsuka, T., Araki, S., Ishizuka, K., Fujimoto, M., Heinrich, M., Yamato, J.: A realtime multimodal system for analyzing group meetings by combining face pose tracking and speaker diarization. In: *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI 2008)*, Chania, Crete, Greece, pp. 257–264 (2008)
16. Ohtsuka, S., Oka, S., Kihara, K., Tsuruda, T., Seki, M.: Human-body swing affects visibility of scrolled characters with direction dependency. In: *Society for Information Display (SID) 2011 Symposium Digest of Technical Papers*, pp. 309–312 (2011)
17. OpenCV. <http://opencv.org/>

18. OpenFace API documentation. <http://cmusatyalab.github.io/openface/>
19. Padmavathi, G., Shanmugapriya, D., Kalaivan, M.: A study on vehicle detection and tracking using. *Wirel. Sens. Netw.* **2**, 173–185 (2010)
20. Schoff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: *IEEE Conference on CVPR 2015*, pp. 815–823 (2015)
21. Sirkin, D., Ju, W.: Consistency in physical and on-screen action improves perceptions of telepresence robots. In: *HRI 2012 Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 57–64 (2012)
22. Tariq, A.M., Ito, T.: Master-slave robotic arm manipulation for communication robot. In: *Japan Society of Mechanical Engineer, Proceedings of 2011 Annual Meeting*, vol. 11, no. 1, p. S12013, September 2011
23. Ubuntu. <https://www.ubuntu.com/>
24. Watanabe, T.: Human-entrained embodied interaction and communication technology. In: *Fukuda, S. (ed.) Emotional Engineering*, pp. 161–177. Springer, London (2011). https://doi.org/10.1007/978-1-84996-423-4_9
25. Wongphati, M., Matsuda, Y., Osawa, H., Imai, M.: Where do you want to use a robotic arm ? And what do you want from the robot ? In: *International Symposium on Robot and Human Interactive Communication*, pp. 322–327 (2012)