Deep Level Situation Understanding and its Application to Casual Communication between Robots and Humans

Yongkang Tang[1], Fangyan Dong[1], Mina Yuhki[1], Yoichi Yamazaki[2],
Takanori Shibata[3] and Kaoru Hirota[1]

[1]*Dept. of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering,*
Tokyo Institute of Technology, G3-49, 4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan

[2]*Dept. of Electrical, Electronic & Information Engineering, Faculty of Engineering, Kanto Gauin University,*
1-50-1 Mutsuura-higashi, Kanazawa-ku, Yokohama 236-8501, Japan

[3]*Human Technology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST),*
Central 6, 1-1-1Higashi, Tsukuba 305-8566, Japan

Keywords: Human Robot Interaction, Agent, Emotion Understanding, PARO.

Abstract: The concept of Deep Level Situation Understanding is proposed to realize human-like natural communication among agents (e.g., humans and robots/machines), where it consists of surface level understanding (such as gesture/posture recognition, facial expression recognition, and speech/voice recognition), emotion understanding, intention understanding, and atmosphere understanding by applying customised knowledge of each agent and by taking considerations to careful attentions. It aims to not impose burden on humans in human-machine communication, to realize harmonious communication by excluding unnecessary troubles or misunderstandings among agents, and finally to create a peaceful, happy, and prosperous humans-robots society. A scenario is established to demonstrate several communication activities between a businessman and a secretary-robot/a human-boss/a waitress-robot/a human-partner/a therapy-robot (PARO) in one day.

1 INTRODUCTION

Robots are increasingly capable of co-existing with humans in environment, such as in manufacturing, offices, restaurants, hospitals, elder care facilities, and homes. The ability of comprehending human activities, e.g., gesture/posture, speech, and emotion, is required for robots in casual communication, i.e., human to human like communication. Verbal and non-verbal communications are the two basic ways casual communications transmit among various agents such as humans and robots/machines. Several Spoken Dialog Systems are proposed for verbal communication (Sasajima et al., 1999) (Jung et al., 2009). As for nonverbal approaches, gesture recognition has become an attractive research theme in the field of Human-Robot Interaction (Shan et al. 2007), sign language recognition (Vinciarelli et al., 2008. Most works on gesture recognition for Human Robot Interaction (HRI) have been done based on visual information. To improve the robustness of gesture recognition system, a Choquet integral based multimodal gesture recognition system (Tang et al., 2011) is proposed.

In casual communication among humans, a human may hide their real emotions, intentions, and opinions. But other humans may be able to understand them to some extent by understanding the spoken contents, voice tones, and facial expression changes. These kinds of audible information (e.g., speech and voice) and visible information (e.g., gesture, posture, and facial expression) are called surface level communication in this paper, while deep level situation understanding is characterized by unifying the surface level understanding, emotion understanding, intention understanding, and atmosphere understanding by applying careful attention to both universal and agent dependent customized knowledge. The deep level situation understanding framework consists of a gesture/posture recognition module, speech/voice recognition module, emotion recognition module, intention understanding module, atmosphere understanding module, and

Tang Y., Dong F., Yuhki M., Yamazaki Y., Shibata T. and Hirota K..
Deep Level Situation Understanding and its Application to Casual Communication between Robots and Humans.
DOI: 10.5220/0004480402920299
In *Proceedings of the 10th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2013), pages 292-299
ISBN: 978-989-8565-71-6

knowledge base (including universal knowledge and customized agent-dependent knowledge). A situation inference network is employed for inferring appropriate responses (e.g., speech, gesture, and facial expression).

The deep level situation understanding in casual communication among various agents, e.g., humans and robots/machines, aims at three issues. Firstly, humans must pay special attention to robots in the ordinary human-machine communication systems, but such burden may be reduced if robots have deep level situation understanding abilities. Secondly, in the real world, unnecessary troubles or misunderstandings in human to human communications may sometimes happen but the deep level situation understanding can make it possible to avoid such lower level troubles. The customized agent-dependent knowledge will help to comprehend and avoid miscommunication. Thirdly, with the consideration of surface level information, emotions, intentions, atmospheres, universal knowledge, and customized agent-dependent knowledge, it will also help to understand the background, habits, and intention of the agent for smoothing natural Human Robot Interaction, so as to create a peaceful, happy, and prosperous society which is consisted of humans and various specification robots/machines.

To illustrate such a peaceful, happy, and prosperous humans-robots society, a short story is demonstrated by four humans, two eye robots, and a therapy-robot PARO.

Surface level understanding is summarized in 2. In 3, concept of deep level situation understanding is proposed. A scenario is demonstrated to illustrate the proposed deep level understanding in 4.

2 SURFACE LEVEL, EMOTION INTENTION, AND ATMOSPHERE UNDERSTANDING FOR HUMANS-ROBOTS INTERACTION

2.1 Surface Level Understanding and Deep Level Situation Understanding

The audible information (e.g., speech and voice) and visible information (e.g., gesture, posture, and facial expression) are just the surface information of humans. Thus the understanding of such surface information is called surface level understanding in this paper. If the understanding level is illustrated by an iceberg (Figure 1), the audible and visible information is just like a tip above the sea level of the whole iceberg, while there still remain more information hidden under the sea level such as emotion, intention, and atmosphere. In contrast with surface level understanding, the deep level situation understanding is characterized by unifying the surface level understanding, emotion understanding, intention understanding, and atmosphere understanding by adding a careful attention function to the inference engine on the situation network consisted of both universal knowledge and agent dependent customized knowledge. The relationship below the surface level understanding and the deep level situation understanding is illustrated in Figure 1.

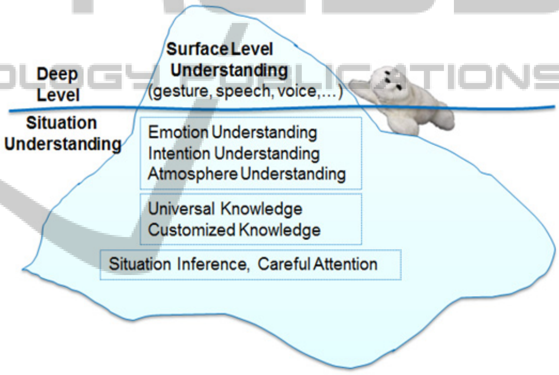

Figure 1: The relationship of the surface level understanding and the deep level situation understanding.

Surface level understanding such as speech understanding and gesture/posture understanding are summarized in 2.2 and 2.3, respectively. Emotion understanding, intention understanding, and atmosphere understanding are summarized in 2.4, 2.5, and 2.6 respectively.

2.2 Speech Understanding

The spoken Dialog System (SDS) provides a communication interface between the user and a computer-based system with a restricted domain in its manner of speech. In the SDS, the dialog is conducted by predefined rules. It is not flexible enough to adapt the natural dialog flow. A Markov Decision Process based dialog system (Levin et al., 2000) is proposed. They argued that a dialog system can be mapped to a Markov Decision Process with additional assumption about the state transition probabilities. A reinforcement learning algorithm is employed to find the optimal strategy. The Partially Observable Markov Decision Processes for dialogue modelling (Williams et al., 2008) is proposed by extending Markov Decision Processes, providing a principled account of noisy observations. The result outperforms that of Markov Decision Processes based method.

2.3 Gesture/Posture Understanding

Computer vision based human gesture recognition systems often get motion information of body parts (hands and head) by applying a skin colour tracking method. It is, however, easy to have noise caused by objects which have similar colour to the skin. On the other hand, application of accelerometer data to gesture recognition is an emerging technique to improve recognition performance. An air writing system is proposed (Amma et al., 2010), which recognizes the gestures of writing alphabets by using wearable acceleration sensors.

A multimodal gesture recognition system is proposed in (Yamazaki et al., 2010), where information of both 3D acceleration and images are combined based on fuzzy logic. In their study, when the similarity calculated from the acceleration recognition units is greater than a given threshold value, the result from the accelerometer is taken as the final result. In reverse, the image recognition unit processes the candidate gestures that come from the acceleration unit to get the final result. How to decide the given threshold, however, should be investigated to apply the method for other cases of gestures in general.

To deal with these problems and to be able to apply to fusing two units, a Choquet integral based multimodal gesture recognition method is proposed (Tang et al., 2011), where human gestures are recognized based on the fusion of video images and 3D acceleration sensors. By calculating the optimal fuzzy measures of the Camera-based recognition unit and the accelerometer-based recognition unit, the gesture recognition system achieves a recognition rate over 92% for eight types of gestures.

2.4 Emotion Understanding

An automatic real-time capable continual facial expression recognition system is proposed (Hommel and Handmann, 2011) based on Active Appearance Models (AAMs) and Support Vector Machines (SVMs) where face images are categorized into seven emotion states (neutral, happy, sad, disgust, surprise, fear, and anger). An individual mean face is estimated over time to reduce the influence of individual features. Biosensor is also used for emotion understanding. An online affect detection system is proposed (Liu et al., 2008) by using wearable biofeedback sensors. Physiological signals (temperature, pulse cycle etc.) acquired form the sensors are preprocessed and then classified by Support Vector Machine (SVM).

In casual communication, emotion may be expressed in both facial expression and voice. A multimodal emotion recognition system is proposed (Paleari et al., 2010) to recognize emotions from audio sequence and static images.

2.5 Intention Understanding

Estimating the intention of human is also important in Human-Robot Interaction. An intention reasoning algorithm (Takagi et al., 2000) is proposed based on a bidirectional associative memories driven support system. A maximum entropy based intention understanding method (Shimada et al., 2007) is proposed for understanding the intention of speech in a dialog system.

2.6 Atmosphere Understanding

In communication among multiple agents, e.g., a conference with twenty participants, it may not be easy to identify the attitude, mood, and emotion of each individual. Instead the atmosphere of the entire gathering becomes an important issue for smooth communication. Several attempts have been made to define the communication atmosphere. A 3D coordinates with physical dimension, social dimension, and mental dimension are built to represent the atmosphere in human communications (Rutkowski et al., 2004). To reflect the uncertainty and subjectivity of the atmosphere as well as its effect on the emotions of the individuals in many-to-many communication, a concept of Fuzzy Atmosfield (FA) is proposed (Liu et al., 2011) to represent the atmosphere being created in the process of interactive communication.

To adapt robot's behaviour for smooth communication in human robot interaction, a fuzzy production rule based friend-Q learning method (FPRFQ) is introduced (Chen et al., 2012). Based on the FPRFQ, a behaviour adaptation mechanism is proposed to solve the robots' behaviour adaptation problem.

3 CASUAL COMMUNICATION BASED ON DEEP LEVEL SITUATION UNDERSTANDING

3.1 Deep Level Situation Understanding

Although speech understanding, gesture/posture understanding, emotion understanding, intention understanding, and atmosphere understanding can help robot to comprehend parts of human activities, it is still not sufficient for casual Human Robot Interaction. People usually hide their real emotions, intentions, and opinions and show them in another indirect/different way. These kinds of information are just a reflection of the real emotions, intentions, and opinions.

The deep level situation understanding is characterized in 2.1. There are many instances of deep level situation understanding in daily life. For example, when you need someone's help, you may ask "Are you busy?" rather than "May I ask you for a favour?" because of manners. Usually the person you are talking to will know that you need his/her help. Suppose you visit a convenience store and want to buy a fountain pen. In this case, you may ask the shop assistant that "Do you have a fountain pen?" The shop assistant will know that you want to buy this kind of pen. Even if it is a yes-no question, neither "yes" nor "no" is expected to end the conversation. If there are fountain pens in the shop, the shop assistant will guide the customer to the specific location of the fountain pens. If they do not have this kind of pen, in order to provide satisfactory service to the customer, they will guide the customer to the shop where the fountain pen can be purchased. Imagine a lady usually goes to a cafe for her favourite coffee and dessert. The waiter/waitress in the cafe knows the preference of their regular customer. When this lady just orders "the usual one" it is no doubt that the waiter/waitress will understand the meaning and bring the desired drink and dessert to her.

3.2 Situation Network

3.2.1 Multimodal Framework for Deep Level Situation Understanding

The multi-modal framework for deep level situation understanding is shown in Figure 2. Speech content, emotion, gesture/posture, Speech information recorded with the microphone is processed into speech content by speech recognition system. Meanwhile, emotion state can be estimated from acoustic features of the speaker. Sometimes emotion may be expressed by facial expressions. Emotion expressed on the face can be recognized by facial expression based emotion recognition algorithm (e.g., Hommel et al., 2011). Atmosphere can be estimated based on the emotion state of the agents. Atmosphere information also can be estimated from the emotion changes in the conversation. Gestures/Postures can be recognized by gesture recognition algorithms (e.g., Tang et al., 2011) from sensors like cameras and accelerometers.

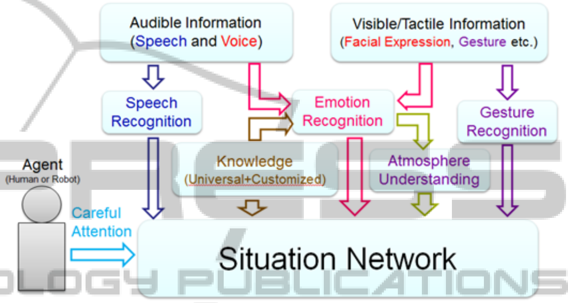

Figure 2: Framework for Deep Level Situation Understanding.

3.2.2 Situation Network

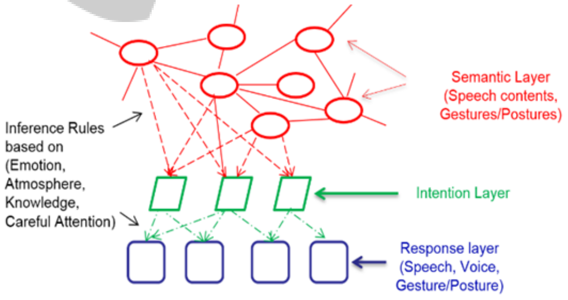
Semantic Layer
(Speech contents, Gestures/Postures)
Inference Rules based on (Emotion, Atmosphere, Knowledge, Careful Attention)
Intention Layer
Response layer
(Speech, Voice, Gesture/Posture)

Figure 3: Situation network.

Careful attention from agents, speech content, knowledge (including universal knowledge and customized knowledge), emotion, atmosphere, and gesture are employed by the situation network for reasoning an appreciated response to the communication situation. The situation network consists of a semantic layer, an intention layer, and an expression layer, as shown in Figure 3. The semantic layer is composed of nodes and links. The nodes in the semantic layer denote objects. The edges between two nodes are the relationship of nodes. The semantic layer can be established with universal and customized agent-dependent knowledge in advance. The information achieved from the communication (e.g., speech content and gestures/postures) are employed to infer the key nodes in the semantic layer.

The intention of the interaction can be inferred from the key nodes of the semantic layer to the intention layer by referring to information like current emotion state, atmosphere, and knowledge (including universal knowledge and customized agent-dependent knowledge). Finally with the comprehensive consideration of the current emotion state, atmosphere, universal and customized agent-dependent knowledge, and careful attention, suitable response (speech, voice, and gesture/posture) will be reasoned as the final reaction.

4 DEMONSTRATION SCENARIO ON "ONE DAY OF A BUSINESS MAN"

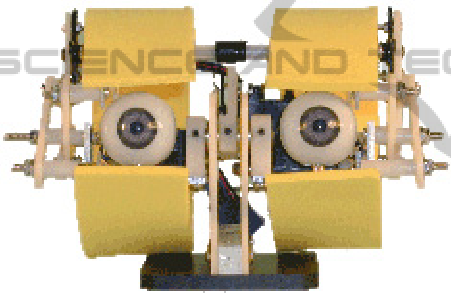

Figure 4: Eye robot.

Figure 5: PARO robot.

To illustrate the applicability of the proposed deep level situation understanding mechanism, a demonstration scenario, entitled "One day of a businessman", is established to narrate several communication activities between a businessman and secretary-robot/human-boss/waitress-robot/ human-partner/therapy-robot (PARO) in one day. Six episodes of deep level situation communication are created with comprehensive consideration of speech content, gesture, emotion, intention, atmosphere, universal knowledge and customized agent-dependent knowledge, and careful attention function. In the scenario, the businessman is asked to reserve a meeting room with a secretary-robot for remote TV meeting. After reporting to his boss, he notices that he made a mistake reserving the meeting. He asks the secretary-robot to help him to change the meeting schedule. After one day's hard work, he goes to a small Japanese-style restaurant for his favourite food. In both cases that a businessman goes to the restaurant as a new customer and the case that the businessman goes to the restaurant as a regular customer are demonstrated for comparing the surface level communication and deep level communication. The communication between the businessman and his wife is narrated in the last scene 6.

The scenario is demonstrated by four humans, two eye robots (Figure 4), and a therapy-robot PARO (Figure 5).The script of the scenario is shown in Table 1-6 and the scenes are shown in Figure 6-11.

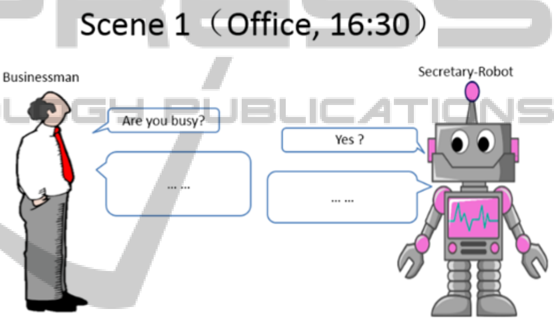

Figure 6: Reserving the meeting room.

Table 1: The businessman talks with a secretary-robot to reserve a meeting room. (*E, I, A, K, and C* stand for Emotion, Intention, Atmosphere, customized Knowledge, and Careful attention, respectively).

Place: Office, Time: 16:30
Businessman: Are you busy now?
Secretary-robot: May I help you (*I*)?
Businessman: Is the meeting room available after 3 PM this Thursday?
Secretary-robot: It could be reserved (*I*) after 3:30 pm.
Businessman: Great (*I*)! A quiet meeting room will be better.
Secretary-robot: Yes, it's the meeting room on the 17th floor as always (*K*).
Businessman: Fine (*I*). Please reserve it till 5 pm, as we'll have a meeting with our branch office.
Secretary-robot: No problem. I'll also reserve the TV conference system (*I, C*) for you.
Businessman: Thank you so much!
Robot: You're welcome.

Table 2: The businessman reports to his boss and found the meeting date he reserved was wrong.

Place: Office, Time: 16:50
Businessman: Boss, is it OK now?
Boss: What's up? (*I*)
Businessman: I have reserved the meeting room and TV system for the meeting with the branch office.
Boss: Great! Good job (*C*)! It's the next Thursday, right?
Businessman: Ah, I've got it wrong. I reserved it on this Thursday.
Boss: Oh, not again! (*K*)
Businessman: I'm going to change it right away (*I, A*). I apologize for that.
Boss: Be careful this time.(*I*)
Businessman: Yes, Sir. I'm so sorry. I'm leaving now.

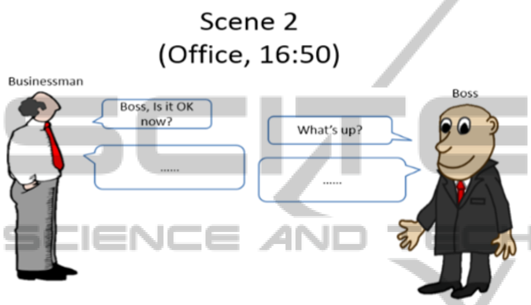

Figure 7: Reporting to boss.

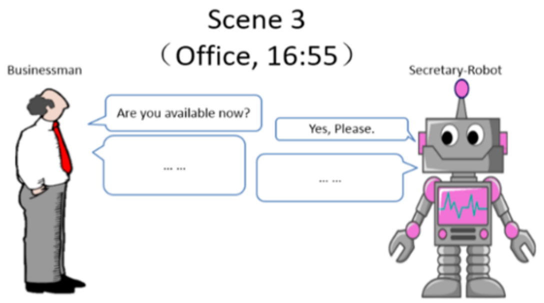

Figure 8: Changing the meeting date.

Table 3: The businessman returns to the robot to change the reservation date.

Place: Office, Time: 16:55
Businessman: Are you available now?
Robot: Yes, please (*I, A*).
Businessman: Sorry (*C*), may I change the reservation of the meeting with branch office to next Thursday?
Robot: Got it! I'm checking (*I*) the status of meeting room right now (*E*).
Businessman: Great. Please help to check (*C*).
Robot: All of the meeting rooms on the 17th floor have been reserved on next Thursday. How about meeting room on the 11th floor (*K*)?
Businessman: Thanks a lot! You really saved me!
Robot: The reservation of the TV conference system also needs to be changed (*I, C*), right?
Businessman: Sure. Thank you as always (*C*)!
Robot: You're welcome!

Table 4: A waitress-robot entertains an ordinary customer in a small Japanese-style restaurant.

Place: Small Japanese-style restaurant, Time: 18:00
Waitress-robot: Hi, may I help you?
Businessman: [abnormal face] Hi.
Waitress-robot: Which seat do you prefer?
Businessman: Next to the window.
Waitress-robot: May I take your order now?
Businessman: Yes, I want to have a stewed fish, please.
Waitress-robot: Would you like something to drink?
Businessman: Draft beer.
Waitress-robot: Just a moment please, Sir.
Businessman: All right.

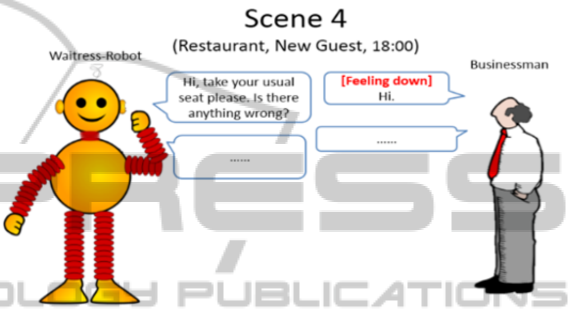
Scene 4
(Restaurant, New Guest, 18:00)
Waitress-Robot
Businessman
Hi, take your usual seat please. Is there anything wrong?
[Feeling down] Hi.
......
......

Figure 9: At the restaurant as a new customer.

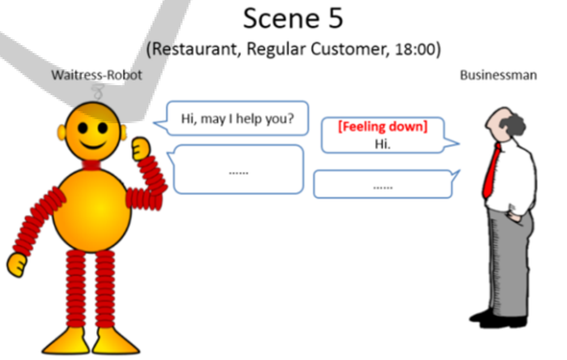

Figure 10: At the restaurant as a regular customer.

Table 5: A robot entertains a regular customer in a small Japanese-style restaurant.

Place: Small Japanese-style restaurant, Time: 18:00
Businessman: [abnormal face] Hi!
Waitress-robot: Hi. Take your usual seat (*K*) please. Is there anything wrong (*E, C, K, A*)?
Businessman: Yeah, a little bit.
Waitress-robot: Let forget the unpleasantness with delicious cooking!
Businessman: Thank you (*E, C*).
Waitress-robot: Is the usual (*K*) order, OK?
Businessman: Yes.
Waitress-robot: Draft beer (*K*) for drinks, is OK?
Businessman: OK. Thanks (*C*)!
Waitress-robot: Green soybeans are also served for free (*E, C, K*).
Businessman: Wonderful! Thank you!

Table 6: The businessman goes back home and talks with his wife.

Place: Home, Time: 21:00
Businessman: [feeling down voice] I'm home!
Wife: Welcome back! Are you drinking (*K*)?
Businessman: Well…
PARO: (Joyful) sound as owner is back.
Businessman: Yeah. Hello, PARO.
PARO: (Joyful) sound as respond to owner's greeting.
Wife: Would you like to take a bath or a cup of tea?
Businessman: Well…
Wife: What happened (*E, A, K*)?
Businessman: Well…
Wife: Made a mistake (*A, K*)?
Businessman: Yes…
Wife: "Hana" was waiting for you!
Businessman: For playing game together (*I, K*)?
Wife: Yes.
Businessman: Tomorrow (*K*), I'll play with her. I'm a little bit tired.
Wife: OK, I will prepare the bath (*I, C*).
Businessman: Well, thanks.

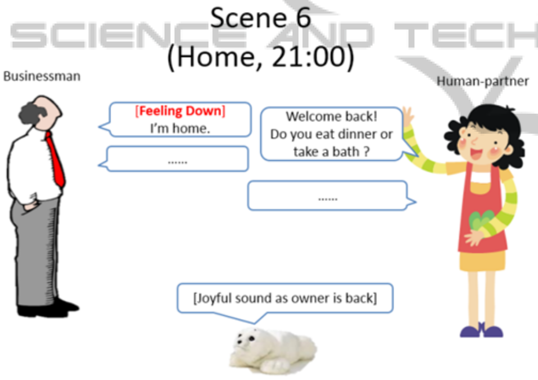
Scene 6
(Home, 21:00)
Businessman
Human-partner
[Feeling Down] I'm home.
……
Welcome back! Do you eat dinner or take a bath ?
……
[Joyful sound as owner is back]

Figure 11: At home.

A surface level understanding between a human and a robot is shown in scene 4; while a deep level situation understanding is narrated in scene 5. Because the waitress-robot in the scene 5 knows the face of the businessman, she is aware that something bad happened to the businessman. Because he is a regular customer to this restaurant, the waitress-robot know his taste well and offers comfortable service to the businessman. Based on the abnormal facial expression, the waitress-robot notice that the regular customer is a little sad. So she serves a friend-like service to warm the customer.

5 CONCLUSIONS

A story of "one day of a business man" demonstrates the deep level situation understanding in casual communications among humans and robots/machines. As shown in scene 3, when the business man tells the secretary-robot that he made a mistake on the meeting date, the robot estimates his intention of re-reserving the meeting room; the robot also knows he is in a hurry to change it. So the secretary-robot responses by "checking the meeting room immediately" to calm the businessman down. After changing the date, the secretary-robot also informs the businessman to change the reservation of the TV conference system to avoid mistake. This kind of high level communication may be impossible in surface level communications.

Not only surface level understanding (e.g., speech/voice recognition, gesture/posture recognition), emotion understanding, intention understanding, and atmosphere understanding but also customized agent-dependent and universal knowledge, and a careful attention mechanism are considered for smoothing and naturalizing communication among humans and robots/machines. The proposal aims at making the robot have high level human-like communication abilities. This will reduce the burden on human when interacting with robots/machines. By considering the customized agent-dependent knowledge in human-robot communication, it will help robots/machines to understand the usual way in communication and avoid unnecessary troubles and misunderstandings. With the comprehensive consideration of speech/voice, gesture/posture, emotion, intention, atmosphere, and knowledge (including universal and customized knowledge), the proposal will smooth the communication between humans and robots/machines as well as create a peaceful, pleasant, and prosperous society consisting of humans and various specification robots.

More and more, robots are required to do house work, care the elder, look after children, and work in the office. Hence the ability to communicate with persons at all ages is becoming essential. The proposal can smooth human robot interactions by considering necessary factors (e.g., customized agent-dependent knowledge) in communications. Furthermore the proposed deep level situation understanding may help to build the coexistence and co-prosperity in the human-robot society in the near future.

ACKNOWLEDGEMENTS

This research project is supported by Japan Society for the Promotion of Science (No.21300080).

REFERENCES

Amma, C., Gehrig, D., & Schultz, T., 2010. Airwriting recognition using wearable motion sensors. In *Proceedings of the 1st Augmented Human International Conference*. ACM.

Chen, L., Liu, Z. et al., 2012. Multi-Robot Behavior Adaptation to Communication Atmosphere in Humans-Robots Interaction Using Fuzzy Production Rule Based Friend-Q learning, *International Symposium on Soft Computing*.

Hommel, S., & Handmann, U., 2011. AAM based continuous facial expression recognition for face image sequences. In *Computational Intelligence and Informatics (CINTI), 2011 IEEE 12th International Symposium*, 189-194.

Jung, S., Lee, C., Kim, K., Jeong, M., & Lee, G. G., 2009. Data-driven user simulation for automated evaluation of spoken dialog systems. *Computer Speech & Language*, 23(4), 479-509.

Levin, E., Pieraccini, R., & Eckert, W., 2000. A stochastic model of human-machine interaction for learning dialog strategies. *Speech and Audio Processing, IEEE Transactions*, 8(1), 11-23.

Liu, C., Conn, K., Sarkar, N., & Stone, W., 2008. Online affect detection and robot behavior adaptation for intervention of children with autism. Robotics, *IEEE Transactions on Robotics*, 24(4), 883-896.

Liu, Z. T., Dong, F. Y., Hirota, K., Wu, M., Li, D. Y., & Yamazaki, Y., 2011. Emotional states based 3-D Fuzzy Atmosfield for casual communication between humans and robots. In *Fuzzy Systems (FUZZ), 2011 IEEE International Conference,* 777-782.

Paleari, M., Huet, B., & Chellali, R., 2010. Towards multimodal emotion recognition: a new approach. In *Proceedings of the ACM International Conference on Image and Video Retrieval,* 174-181.

Rutkowski, T. M., Kakusho, K., Kryssanov, V., & Minoh, M., 2004. Evaluation of the communication atmosphere. In *Knowledge-based intelligent information and engineering systems,* 364-370.

Sasajima, M., Yano, T., & Kono, Y., 1999. EUROPA: A generic framework for developing spoken dialogue systems. In *Proc. of EUROSPEECH'99,* 1163-1166

Shan, C., Tan, T., & Wei, Y., 2007. Real-time hand tracking using a mean shift embedded particle filter. *Pattern Recognition*, 40(7), 1958-1970.

Shimada, K., Iwashita, K., & Endo, T., 2007. A case study of comparison of several methods for corpus-based speech intention identification. In *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics,* 255-262.

Takagi, T., Nishi, T., & Yasuda, D., 2000. Computer assisted driving support based on intention reasoning. In *Industrial Electronics Society, 2000. IECON 2000. 26th Annual Confjerence of the IEEE,* (1), 505-508.

Tang, Y., Hai, V. et al., 2011. Multimodal Gesture Recognition for Mascot Robot System Based on Choquet Integral Using Camera and 3D Accelerometers Fusion. *Journal of Advanced Computational Intelligence and Intelligent Informatics*. (15), 563-572

Vinciarelli, A., Pantic, M., Bourlard, H., & Pentland, A., 2008. Social signal processing: state-of-the-art and future perspectives of an emerging domain. In *Proceedings of the 16th ACM international conference on Multimedia*, 1061-1070.

Williams, J. D., Poupart, P., & Young, S., 2008. Partially observable Markov decision processes with continuous observations for dialogue management. In *Recent Trends in Discourse and Dialogue*, 191-217.

Yamazaki, Y., Vu, H et al., 2010. Gesture recognition using combination of acceleration sensor and images for casual communication between robots and humans. In *Evolutionary Computation,* 1-7.