Investigating Appropriate Relative Position of User and AR Character Agent for Communication Using AR WoZ System

Reina Aramaki[1] Makoto Murakami[1]

[1] Department of Information Sciences and Arts, Toyo University
2100 Kujirai, Kawagoe, Saitama, 350–8585 Japan
Email: s36d01300014@toyo.jp, murakami_m@toyo.jp

Abstract: The aim of this study is to construct a system that can communicate with humans in their daily life environments, such as their home or a pedestrian street. We propose a human–agent communication system using augmented reality (AR) technology, in which the AR character agent has no physical body, thus facilitating its safe performance in the real environment of daily life. As the first step in implementing the motion control component of the AR agent, we focused on investigating the appropriate spatial relationship between the user and the AR agent using some experimental settings. We constructed an AR agent Wizard of Oz (WoZ) system, in which the agent is operated by a hidden operator using remote control, to acquire human–agent interaction data through experimental trials. We collected the interaction data in a simple experimental setting: the user sits at a desk and communicates with the AR agent standing in various positions on the desk. We investigated the spatial relationship between the user and the AR agent appropriate for communication. The results show that many subjects felt that it was appropriate to talk with the AR agent when it was straight in front of and approximately 70–93 cm away from them.

1 Introduction

Researchers have tried to develop robots that can cooperate with humans in their daily life. Domestic service robots may touch humans, and should be designed to ensure human safety. Wyrobek et al. [1] developed a mobile 2-armed robot that has a low risk of causing serious injury as compared to former industrial robots. Sugaiwa et al. [2] developed a robot with a shock absorbing skin, which achieves collision safety. In the RoboCup@Home league competition, which consists of benchmark tests to evaluate the abilities of domestic service robots in realistic home environment settings, Stuckler et al. [3] developed a robot with robust navigation, object manipulation, and intuitive interaction abilities that are necessary for domestic service. The Tsukuba Challenge, which is a real-world robot challenge for autonomous mobile robots working in a real public street with pedestrians, has been held in the city center of Tsukuba, Japan, every year since 2007. In 2010, more than 60 robots participated and 7 robots succeeded in the task set [4]. As seen above, many researchers have been developing robots that can safely work with humans in their daily life. However, more time will be needed to develop robots that can be put to practical use.

On the other hand, a communication robot that is not required to do heavy physical work can be designed to be smaller, lighter, and safer than a domestic-service or elder-care robot. However, it is uncomfortable for a standing human to have to look down at a small robot on the floor when communicating with it. Therefore, it is assumed that the small robot is placed on a table or desk. However, given the current technology, it is difficult for the robot to get on and off the desk. Further, in a street, there is nothing for the robot to stand on, and the problem remains.

The problems described above stem from the fact that the robot has a physical body. Only to construct a communication system, by eliminating the physical body, the physical restrictions and related problems can be reduced. Therefore, we propose a human–agent communication system that uses augmented reality (AR) technology in which the AR character agent has no physical body. The user of this system wears a head-mounted display with a camera, a headset microphone, and earphones, as shown in the Fig. 1 and Fig. 2. First, the system captures the user's view with the camera and displays it on the head-mounted display. Then, the system integrates a computer-generated character agent with the user's view using augmented reality technology. The user can communicate with the AR agent using spoken language through the headset microphone and earphones. In this way, users feel that the AR agent is actually present and is communicating with them. In addition, the AR agent can walk around safely in the environment comprising various obstacles, easily get on and off a table or desk, and communicate with the user while it is floating in the air. Therefore, the system is usable in the real environments of daily life, such as at home or on a pedestrian street.

The objective of the study described in this paper was to construct the motion control component of the

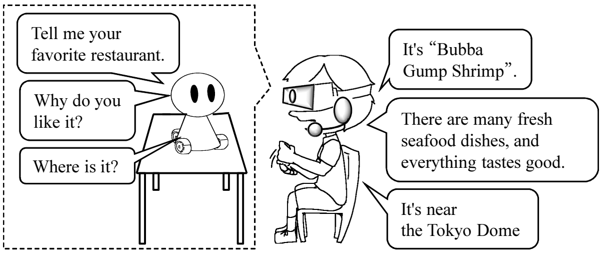

Figure 1: Example of home use of the human–agent communication system using augmented reality technology.

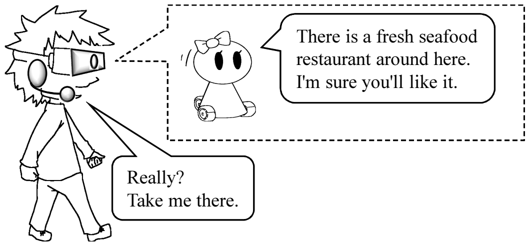

Figure 2: Example of street use of the human–agent communication system using augmented reality technology.

AR agent for the proposed system. In Human–Robot Interaction (HRI) studies, some researchers used the Wizard of Oz (WoZ) method in which the robot is operated by a hidden operator using remote control, and conducted experiments using subjects in various situational or task scenarios to obtain HRI data. In addition, they investigated the appropriate robot motion for the given situations or tasks, and constructed the motion control component of the robot. We constructed an AR agent WoZ system, and as the first step in constructing the motion control component of the AR agent, we investigated the user–AR agent spatial distance and orientation that is appropriate for communication using some experimental settings with an AR WoZ system.

This paper is organized into six sections. After the Introduction, Section 2 presents a human–agent communication system using AR technology. In Section 3, we briefly discuss related work. In Section 4, the AR WoZ system is presented in detail. In Section 5, the experiment is described. Finally, in Section 6, we present our conclusions and discuss future works.

2 Human–Agent Communication System using Augmented Reality Technology

In this section, we describe the implemented human–agent communication system platform. A user wears a head-mounted display Vuzix Wrap1200VR equipped with an ASUS Xtion PRO LIVE, a headset microphone, and earphones. The system captures color image data with an RGB sensor in the Xtion using OpenNI, and displays it on the head-mounted display using OpenGL. The system loads the body model data of the character agent created with the 3D computer graphics (CG) modeling software, and integrates the CG character agent with the user's view on the head-mounted display using OpenGL. The head-mounted display Wrap1200VR has a 3-degree-of-freedom (DoF) head tracker, and the system can display the AR character agent in the correct position according to the user's head direction. Fig. 3 shows the user's view displayed on the head-mounted display when the user looks forward and then looks to the right. The Xtion has a depth sensor, and the system can determine whether objects in the real world are behind the AR character agent. The system can display the agent as if it were behind the real objects. Fig. 4 shows the user's view when the agent is behind a computer display or the user's hand. We implemented the speech recognition component using Julius, which is a large-vocabulary continuous speech recognition decoder [5], and the speech synthesis component using Hoya VoiceText Engine SDK.

We did not implement the motion control component and a dialogue management component of the AR character agent, which make the AR agent behave intelligently, because the appropriate motion and dialogue strategy depends on the situation or task. In this study, as a first step in implementing the AR agent's motion control component, we focused on investigating the appropriate user–AR agent spatial distance and orientation using some experimental settings.

Figure 3: User's view displayed in the head-mounted display when the user looks forward and then looks to the right.

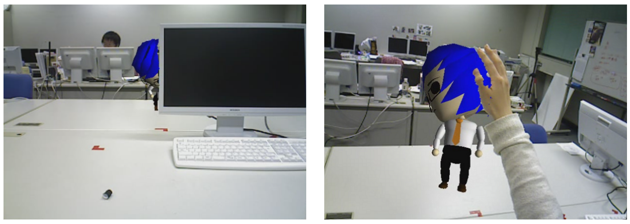

Figure 4: User's view displayed in the head-mounted display when the AR agent is behind real objects.

3 Related Work

Hall [6] studied interpersonal distance, and coined the term proxemics. He categorized interpersonal distance into four classes: intimate, personal, social, and public. At an intimate distance, which is closer than 45 cm, lovers or close friends communicate with each other. At a personal distance, which is from 45 cm to 1.2 m, friends communicate with each other. At a social distance, which is from 1.2 m to 3.6 m, non-friends or strangers have a conversation with each other. The public distance, which is greater than 3.6 m, is used for making a public speech.

When two or more people communicate with each other, their spatial and orientational relationship is established and sustained. Kendon [7] called such a system of behavioral organization an F-formation system. For example, a circular arrangement appears when three or more persons are standing in free space and have a conversation. Examples of the F-formation of two persons are a vis-à-vis arrangement in which they stand and face each other directly, an L-arrangement in which they stand such that they orthogonalize each other's gaze direction, and a side-by-side arrangement in which they stand facing in the same direction.

Some researchers in the field of HRI have investigated the appropriate relative position of a robot to a human when the robot is interacting with the human. Walters et al. [8] used a robot that is mechanistic in appearance and 1.1 m tall to perform experimental trials using the WoZ method, and investigated the interpersonal distance comfortable for a human subject when the subject approaches the robot and when the robot approaches the subject. The results showed that approximately 40% of subjects approached the robot and allowed the robot to approach them within the intimate distance defined in Hall's proxemics. Then, they explored the relationship between the preferred approach distance and the subjects' personalities, and found that the "proactiveness" personality factor correlates with the social distance.

Huettenrauch et al. [9] used the same robot as Walters et al. to investigate the spatial relationship between a robot and a subject interacting with it, when the subject shows the robot around and teaches it places and objects in a home-like environment. They categorized the interaction into three events: "FOLLOW" in which the subject guides the robot around, "SHOW" in which the subject teaches the robot places and objects, and "VALIDATE" in which the subject tests the taught places and objects by sending it on missions to find them again. The results show that the Hall's personal distance was the most preferred by the subjects independent of event type, and the number of subjects interacting with the robot at the intimate distance was much smaller than the preferred robot approach distance reported by Walters et al. In addition, they found that the Kendon's vis-à-vis arrangement was the most observed, independent of event type.

Woods et al. [10] investigated how a robot should approach human subjects when performing the fetching and carrying tasks that a domestic robot is expected to execute. They performed an experiment in which the robot approached the subject from various directions in four different situations: the subject seated on a chair in the middle of an open space, standing in the middle of an open space, seated at a table in the middle of an open space, and standing with his/her back against a wall. The results show that the front left and front right approaches were rated as the most comfortable by the subjects in each situation, and when the subjects were standing in an open space, the frontal approach was rated as the most efficient.

Some researchers used immersive virtual environment technology (IVET), and investigated the interpersonal distance maintained between subjects and a virtual human. This approach allows researchers to conduct ecologically realistic experiments with more precise experimental control than does the real environment approach. Bailenson et al. [11] used IVET and investigated the interpersonal distance when a subject approached a stationary standing virtual human, while varying the characteristics of gender, agency (agent vs. avatar; i.e., whether it was apparently controlled by a computer or by another human), and gaze behavior (mutual gaze or not). The average front minimum distance (i.e., the minimum distance when the subject was in front of the virtual human) was 51 cm, and the mean back minimum distance (i.e., when the subject was behind the virtual human) was 45 cm, and the difference between them was significant. The subjects maintained a greater distance from the female virtual human than the male virtual human, and a greater distance from agents who engaged them in a mutual gaze than from agents who did not; this difference did not occur in the case of avatars. The female subjects stayed farther away from avatars than from agents; male subjects did not show this difference. In addition, they measured how far subjects moved away from the virtual human when it walked toward them and invaded their personal space. The results show that the subjects moved farther away from an approaching agent than an approaching avatar.

Llobera et al. [12] conducted an experiment using IVET to examine the impact on subjects' electrodermal activity, which measures the level of physiological arousal, when they were approached by virtual characters. They varied the approached distance (0.4 m, 0.8 m, or 1.6 m), the virtual character's appearance (humanoid or a cylinder of human size), and the number of characters that simultaneously approached (one or four). The results show that the number of skin conductance responses after the approach and the change in skin conductance level increased as the virtual characters approached closer to the subjects. The number of characters was positively associated with the responses, but there was no evidence of a difference in response between the humanoid characters and cylinders.

In our human–agent communication system using AR technology, the AR character agent has no physi-

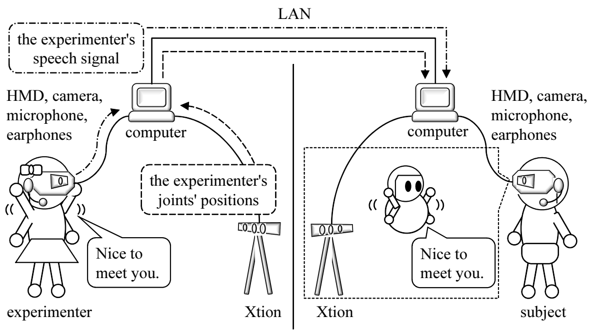

Figure 5: Flow of data from experimenter to subject in the AR WoZ system.

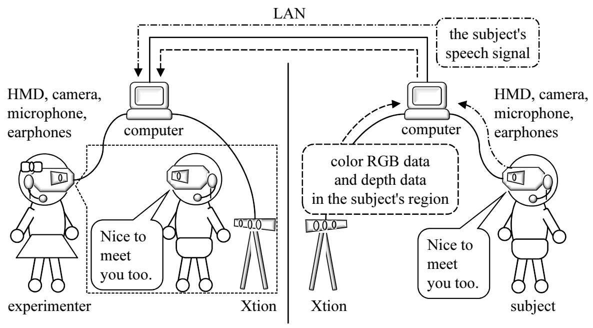

Figure 6: Flow of data from subject to experimenter in the AR WoZ system.

cal body, and can communicate with the user while it is standing on the desk or floating in the air. Therefore, the appropriate relative position of the AR agent to the user can be different from the human–human or human–robot relative position. Our objective is to construct a system that can communicate with users in their daily life, and as the first step in implementing the AR agent's motion control component, we investigated the preferred spatial relationship between the user and the AR agent. Therefore, our study is different from the related studies in which the relative position of human and virtual human was investigated using IVET, which is not used for the real environment.

4 Augmented Reality Wizard of Oz System

In this section, we detail the implemented AR WoZ system.

As shown in Fig. 5 and Fig. 6, an experimenter and a subject were in two different rooms, in each of which there was a computer connected by a LAN. Both the experimenter and subject wore a head-mounted display Vuzix Wrap1200VR with a camera, a headset microphone, and earphones, which were connected to the computer in each room. An ASUS Xtion PRO LIVE was put on the floor or the desk in each room, adequately distant from the experimenter and the subject to capture their whole body or upper body, and connected to each computer.

First, we describe the information presented to the subject. As shown in Fig. 5, the system captures color image data with the camera worn on the subject's head using OpenCV, and displays it in the subject's head-mounted display using OpenGL. The system loads the pre-created body model data of a 3D CG character agent, and displays the character agent overlapping the subject's view in the head-mounted display using OpenGL. On the other hand, with the Xtion in the experimenter's room, the system obtains the experimenter's body joint position data using OpenNI and NITE, and sends it to the computer in the subject's room through the LAN. The system controls the motion of the AR character agent according to the received body joint position data of the experimenter. With a 3-DoF head tracker in the Wrap1200VR, the system can estimate the subject's facial pose, and display the AR character agent in the correct position according to the subject's head direction. The system records the experimenter's speech, sends it to the computer in the subject's room, and outputs it to the subject's earphones. In this way, the subject feels as if the AR agent is actually present and behaves and speaks autonomously.

Now, we describe information presented to the experimenter. As shown in Fig. 6, the system captures color image data with the camera worn on the experimenter's head, and displays it in the experimenter's head-mounted display. On the other hand, with the Xtion in the subject's room, the system obtains color image and depth image data, estimates the subject's body region in the image using OpenNI and NITE, and sends the color and depth data of the subject's body to the computer in the experimenter's room through the LAN. The system calculates 3D polygons of the subject's body from the depth data, sets the color at the polygons vertices using the color data, and displays the colored 3D polygons of the subject's body overlapping the experimenter's view on the head-mounted display using OpenGL. With a head tracker in the head-mounted display, the system can estimate the experimenter's facial pose, and display the subject's body polygon in the correct position, according to the experimenter's head direction. The system records the subject's speech, sends it to the computer in the experimenter's room, and outputs it to the experimenter's earphones. In this way, the experimenter feels as if the subject is actually present and communicating.

Fig. 7 shows the images captured when the AR WoZ system is in operation. The lower images represent the real scene recorded with digital video cameras placed in the rooms of the experimenter and the subject, and the upper images are the experimenter's view and the subject's view displayed on the head-mounted displays in the AR WoZ system.

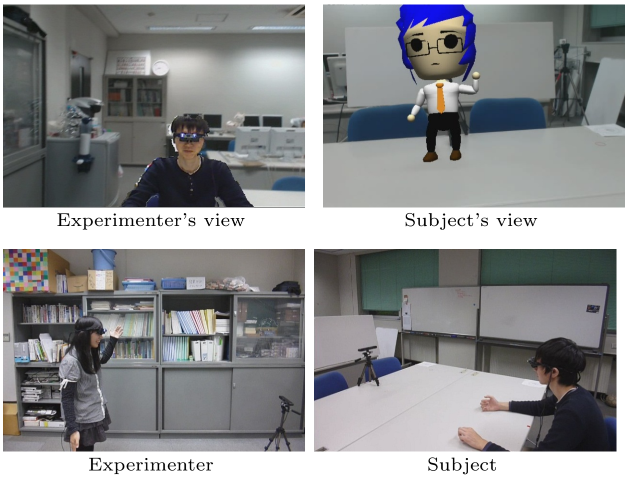

Figure 7: Experimenter's view and subject's view in the AR WoZ system.

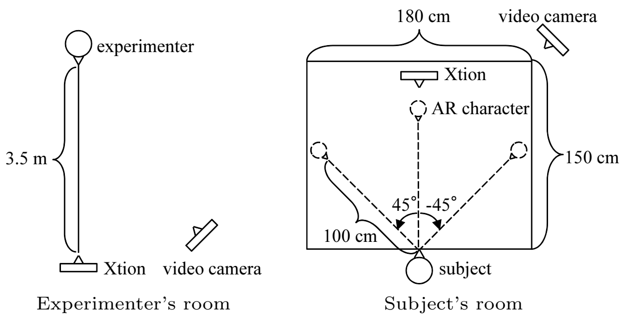

Figure 8: Experimental situation.

5 Experiment

We performed an experiment to explore the appropriate relative position of an AR character agent and a user in a simple experimental setting: the subject sits at a desk and communicates with the AR agent standing in various positions on the desk. The subjects consisted of 7 male and 13 female Japanese university students, and the age of the subjects varied from 21 to 23 years. None of the subjects was paid for participation.

5.1 Experimental Setup and Procedure

The appearance of the AR character agent was as shown in Fig. 7. We set its height as 20 cm. As shown on the right in Fig. 8, the subject sat at an empty 180 cm × 150 cm desk. We put an Xtion on the desk straight in front of the subject and adequately distant to capture the upper body of the subject. As shown on the left in Fig. 8, the experimenter stood in the second room. An Xtion was placed on the floor 3.5 m away from the experimenter. We set up a video camera in each room, as shown in Fig. 8.

We now describe the experimental procedure.

(1) An experiment supervisor introduced and explained the trial procedure to a subject, and the subject signed consent forms and completed introductory questionnaires. Then, the subject put on the head-mounted display with a camera, a headset microphone, and earphones, and sat at the desk.

(2) An experimenter stood 3.5 m straight in front of the Xtion as shown on the left in Fig. 8. The system displayed the AR agent 100 cm away from the subject, straight in front of the subject (0 degrees), in front and diagonally to the left (45 degrees), or in front and diagonally to the right (-45 degrees), as shown on the right in Fig. 8. To familiarize the subject with the system and the situation, the experimenter talked with the subject for 2 min. There were no restrictions on the content of their conversation during these trials.

(3) When the experimenter walked toward the Xtion or stepped away from it, it appeared to the subject that the AR agent walked closer to or stepped away from him/her. Until the subject felt that the distance was appropriate for talking with the AR agent, he/she gave the AR agent (or the hidden experimenter) verbal instructions such as "come closer," "step away," and so on. Then the experimenter talked with the subject for 2 min at the distance that the subject found appropriate for communication.

(4) When the experimenter walked closer to the Xtion, it appeared to the subject that the AR agent walked closer to him/her. When the AR agent was close enough to the subject to allow him/her to talk with it, the subject told it to stop. The subject can give the AR agent further verbal instructions in the same way as in step (3). The experimenter talked with the subject for 2 min at the closest distance that the subject indicated as comfortable for communication.

(5) The experimenter stepped away from the Xtion. The subject gave the AR agent verbal instructions in the same way as in step (4) until the subject felt that the distance was appropriate for talking with the AR agent again. The experimenter talked with the subject for 2 min at the distance that the subject found appropriate for communication.

(6) We repeated steps (3)–(5) three times, changing the direction, as shown on the right in Fig. 8, so that the AR agent was straight in front of the subject (0 degrees), in front and diagonally to the left (45 degrees), or in front and diagonally to the right (-45 degrees). These three trials were performed in random order for each subject to avoid order effects.

(7) Until the subject felt that the position was appropriate for talking with the AR agent, the subject gave the AR agent verbal instructions such as "come closer," "step away," "move to

the right," "move to the left," and so on. The experimenter talked with the subject for 2 min at the position that the subject found appropriate for communication.

We sampled the experimenter's torso position at a rate of 30 Hz with the Xtion, and scaled it according to the ratio of the real space measurement in the experimenter's room and the virtual space measurement displayed in the subject's head mounted display. We calculated the horizontal distance and direction from the subject to the AR agent. To investigate where the subject was looking while talking with the AR agent, we acquired the horizontal direction of the subject's head at a rate of 160 Hz with the head tracker in the head-mounted display. To observe the behavior and speech of the subject and experimenter after the trials, we recorded all the trials using the video camera located in each room, and captured a sequence of images displayed in the head-mounted displays of both the subject and experimenter using a screen-capture software.

After the trials, we asked the subject which direction — "straight in front," "diagonally left," or "diagonally right" — was found to be the most and least appropriate. We measured the subject's personality traits using the Five Factor Personality Questionnaire (FFPQ), one of the Big Five Personality Tests.

5.2 Results

Twelve subjects (60%) answered that the most appropriate direction was "straight in front," four subjects (20%) answered "diagonally left," and four subjects (20%) answered "diagonally right." Three subjects (15%) answered that the least appropriate direction was "straight in front," ten subjects (50%) answered "diagonally left," and seven subjects (35%) answered "diagonally right." All the subjects were right-handed, and there was not clear relationship between the appropriate direction and the handedness. When the subjects were talking with the AR agent present straight in front, diagonally to the left, and diagonally to the right, the average of the horizontal direction of their head was 0.9 degrees, 43.1 degrees, and -43.2 degrees, respectively. These results show that the subjects turned their head to face the AR agent even if it was diagonally to the left or right of them. The average of the appropriate direction acquired in step (7) was 1.3 degrees, the maximum was 9.4 degrees, and the minimum was -12.0 degrees. All the subjects adjusted the AR agent's position so that it was straight in front. The average of the difference between the appropriate direction and the subject's head direction was 2.2 degrees, and we made sure that the subjects faced the AR agent while talking with it.

These results show that many subjects felt that it was appropriate to talk with the AR agent straight in front of them, and even if it was diagonally to the left or right of them, they turned their head to face it.

The average of the appropriate distance acquired in step (3) and step (5) was 88 cm and 83 cm, respectively; there was no significant difference between these distances when we performed t-tests using a level of significance of 0.05. The average of the appropriate distance when the AR agent was straight in front, diagonally to the left, and diagonally to the right of the subjects was 73 cm, 90 cm, and 93 cm, respectively. The average of the appropriate distance acquired in step (7) was 70 cm. These results show that the appropriate distance for communication with the AR agent was approximately 70–93 cm. Significant correlations between distance that each subject found appropriate and the five factors of his/her personal traits were not found.

The average of the closest distance acquired in step (4) when the AR agent was straight in front, diagonally to the left, and diagonally to the right of the subjects was 43 cm, 56 cm, and 59 cm, respectively. Significant correlations between each subject's closest distance and the five factors of his/her personal traits were not found.

These results show that many subjects felt that it was appropriate to talk with the AR agent when it was approximately 70–93 cm away from them, and the closest distance that they preferred for communication was approximately 43 cm.

6 Conclusion

In this paper, we proposed a human–agent communication system using augmented reality technology that is available in humans' daily life. As the first step in implementing the motion control component of the AR agent in the system, we investigated the appropriate spatial distance and direction from the user to the AR agent through a simple experimental setting: the user sitting at the desk communicated with the AR agent standing in various positions on the desk. The results show that many subjects felt that it was appropriate to talk with the AR agent when it was straight in front of them, and, even if it was not straight in front of them, they turned their head to face it. The distance to the AR agent that the subjects found appropriate was approximately 70–93 cm, and the closest distance that they found comfortable for communication was approximately 43 cm.

When an AR agent standing on the desk communicates with a user sitting at the desk, the AR agent position control strategy we should adopt is the following:

(1) The AR agent should be straight in front of and approximately 70–93 cm away from the user, as shown on the area labeled as (1) in the Fig. 9.

(2) If there is something in the area (1), the AR agent should be to the left or right of and approximately 70–93 cm away from the user as shown on the area labeled as (2) in the Fig. 9.

(3) If the area (1) and (2) is occupied by objects, the AR agent can be as close as 43 cm away

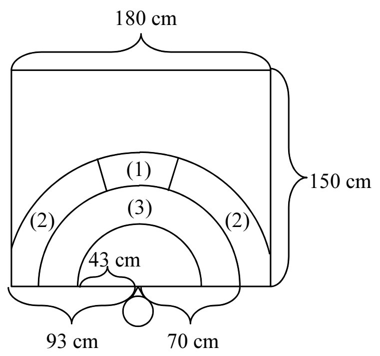

Figure 9: Priority of the area in the desktop where the AR agent should be while talking with the user.

from the user as shown on the area labeled as (3) in the Fig. 9.

In future studies, we will investigate the relative position of the AR agent that users find appropriate using an experimental setting in which something is actually placed somewhere on the desk. The appropriate distance may depend on the size of the AR agent. We will therefore explore the appropriate distance when its size is changed.

References

[1] K. A. Wyrobek, E. H. Berger, H. F. M. der Loos, and J. K. Salisbury, "Towards a personal robotics development platform: Rationale and design of an intrinsically safe personal robot," in *Proceedings of IEEE International Conference on Robotics and Automation*, 2008, pp. 2165–2170.

[2] T. Sugaiwa, H. Iwata, and S. Sugano, "Shock absorbing skin design for human-symbiotic robot at the worst case collision," in *Proceedings of 8th IEEE-RAS International Conference on Humanoid Robots*, 2008, pp. 481–486.

[3] J. Stuckler and S. Behnke, "Integrating indoor mobility, object manipulation, and intuitive interaction for domestic service tasks," in *Proceedings of 9th IEEE-RAS International Conference on Humanoid Robots*, 2009, pp. 506–513.

[4] S. Yuta, M. Mizukawa, H. Hashimoto, H. Tashiro, and T. Okubo, "An open experiment of mobile robot autonomous navigation at the pedestrian streets in the city - Tsukuba Challenge -," in *Proceedings of International Conference on Mechatronics and Automation*, 2011, pp. 904–909.

[5] A. Lee, T. Kawahara, and K. Shikano, "Julius – an open source real-time large vocabulary recognition engine," *Proceedings of EuroSpeech*, pp. 1691–1694, 2001.

[6] T. Hall, Edward, *The Hidden Dimension.* Garden City, N.Y.: Doubleday, 1966.

[7] A. Kendon, *Conducting Interaction: Patterns of Behavior in Focused Encounters.* Cambridge: Cambridge University Press, 1990.

[8] M. Walters, K. Dautenhahn, R. te Boekhorst, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry, "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment," *Proceedings of IEEE International Workshop on Robot and Human Interactive Communication*, pp. 347–352, 2005.

[9] H. Huettenrauch, K. Eklundh, A. Green, and E. Topp, "Investigating Spatial Relationships in Human-Robot Interaction," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5052–5059, Oct. 2006.

[10] S. N. Woods, M. L. Walters, K. L. Koay, and K. Dautenhahn, "Methodological Issues in HRI: A Comparison of Live and Video-Based Methods in Robot to Human Approach Direction Trials," in *Proceedings of 15th IEEE International Symposium on Robot and Human Interactive Communication*, 2006, pp. 51–58.

[11] J. Blascovich, A. C. Beall, J. M. Loomis, J. N. Bailenson, and I. Neuromancer, "Interpersonal Distance in Immersive Virtual Environments," *Personality and Social Psychology Bulletin*, vol. 29, pp. 1–15, 2002.

[12] J. Llobera, B. Spanlang, G. Ruffini, and M. Slater, "Proxemics with multiple dynamic characters in an immersive virtual environment," *ACM Transactions on Applied Perception*, vol. 8, no. 1, pp. 1–12, Oct. 2010.