Merging Viewpoints of User and Avatar in Automatic Control of Avatar-Mediated Communication

Kentaro Ishii[1,2] Yuji Taniguchi[3] Hirotaka Osawa[4] Kazuhiro Nakadai[5] Michita Imai[6,2]

[1] Graduate School Arts and Sciences, The University of Tokyo
[2] Japan Science and Technology Agency, CREST
[3] Graduate School of Science and Technology, Keio University
[4] Faculty of Engineering, Information and Systems, Tsukuba University
[5] Honda Research Institute Japan Co., Ltd.
[6] Faculty of Science and Technology, Keio University

Abstract: This paper discusses the findings on the viewpoint of an avatar-controlling user on the basis of experimentation with an implemented avatar-mediated telecommunication system. Communication using an avatar with facial expressions is useful when a user wants to express emotions. On top of this feature, our system supports automatic avatar movement toward nearest visible location to the target, which is not obvious for the avatar controller. With our system, the avatar controller can easily refer to something remotely. However, sometimes, the words of an avatar controller may not be intuitive for an avatar viewer, because the avatar controller does not necessarily share the viewpoint of the avatar. We designed full-automatic and guide-automatic methods for controlling the avatar, and we conducted an experiment to compare the two methods. The results showed that guide-automatic control was more intuitive than full-automatic control for an avatar viewer, and they have design implications for avatar-mediated telecommunication systems.

1 Introduction

Video phone or chat allows users to visually see the other person and their facial expressions, on top of being able to hear their voice. Communication using such facial expressions is useful when a user wants to express emotions. Meanwhile, techniques for pointing things in remote place have been proposed such as Gesture Laser utilizing an actuated laser pointer [1] and annotation system utilizing visual SLAM [2]. These kinds of techniques help to reduce ambiguity in referencing to something in the real world.

In this paper, we propose an avatar projection system named PROT AVATAR that has both facial expression and remote pointing features. Our prototype employs the PROT system [3] to project the avatar's body onto an arbitrary location like a wall, floor, or ceiling (Figure 1). The face of the controlling user is tracked and captured to create the user's own avatar in remote place. It also has a special sound speaker that emits sounds modulated to ultrasonic wave band. The ultrasonic sound speaker has a feature that it can project a sound source in distant location [4] and we utilize this feature for the avatar viewer to perceive as if the avatar speaks from the projected location. In addition, our prototype supports automatic avatar movement toward nearest visible location of the avatar based on the avatar viewer preference obtained by a preliminary study. As we show later, the avatar viewer preferred seeing the avatar in flat and solid color background when pointing a certain location, even if it was far from the pointed location. Our prototype calculates an appropriate avatar projection location within the preferred area, which is not obvious for the avatar controller. In this way, the system reduces the avatar controller's load to consider an appropriate avatar projected location.

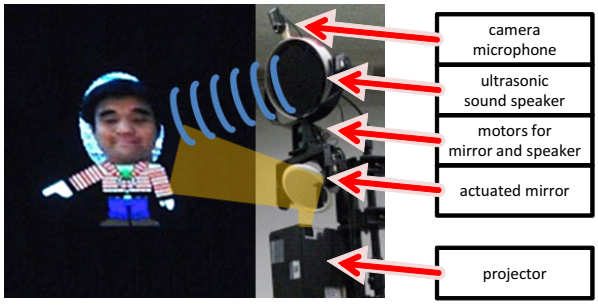


Figure 1: PROT AVATAR, telecommunication system utilizing actuated mirror and ultrasonic speaker for avatar-mediated interaction

However, the avatar controller tends to use unnatural terms for the avatar-viewing user as an interaction partner, because the avatar controller can take a viewpoint different from the avatar viewer's expectation. Since the avatar viewer talks to the avatar, not to the projection system, the avatar viewer naturally expects that the avatar controller takes the avatar's viewpoint. On the other hand, since the

avatar controller is not able to observe the scene from the avatar's viewpoint, the avatar controller tends to take her/his own viewpoint.

We show that careful interaction design can help the avatar controller to adopt the avatar's viewpoint. Imai et al. showed that physical manipulation of an autonomous robot invoke a person to change the person's viewpoint to the robot viewpoint [5]. We explore that similar viewpoint changes occur in the context of controlling virtual avatar. Using the automatically calculated avatar projection location, we designed two control methods: full-automatic control method and guide-automatic control method. With full-automatic control, the avatar is just moved toward the calculated appropriate location when the avatar controller specifies a pointing location, while with guide-automatic control, the calculated appropriate location is first indicated, and the avatar is moved toward that location only after the avatar controller clicks the indicated location. To analyze the avatar controller's viewpoint, we compare the two control methods by investigating usage of demonstrative pronouns, because the interpretation of demonstrative pronouns is sensitive to the viewpoint of the avatar-controlling user. Through an experiment to see the usage of demonstrative pronouns, we discuss the viewpoint changes.

Our contribution is that we propose avatar-mediated telecommunication system that has a function to automatically move avatar toward appropriate location based on a preliminary study to investigate the avatar viewer's preferences, which are not obvious for the avatar controller. In addition, we also perform an experiment to understand users' nature to select demonstrative pronouns and show a method to invoke users to adopt the avatar's viewpoint. We also discuss results and findings for systems using automatic control.

2 Related Work

The avatar can provide remote presence and offer a flexible face-to-face communication environment. Some telecommunication systems or social networking services like Xbox LIVE [6] and Second Life [7] employ avatars. On the other hand, research work like GestureMan [8], Geminoid [9], and field experiment by Shiomi et al. [10] realized remotely controlled robots that can point remote things or provide presence of the robot-controlling users. In addition, Nakanishi et al. showed that camera or display that moves according to an interaction partner enhanced the remote interaction partner's presence during the interaction [11, 12]. In these studies, the scene that the controlling users see changes as the robots or devices move, because the scene is captured by using cameras on them. In contrast, our system employs a projected avatar, which provides faster and more flexible movement of avatar than physical robots. This feature, however, causes the situation that the avatar moves while the remote scene is unchanged, and thus we study the viewpoint of the controlling user in this paper.

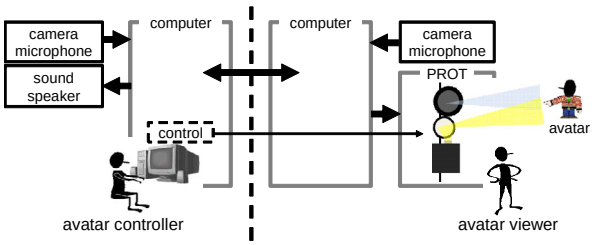


Figure 2: System Configuration. Both avatar controller and viewer side components are network connected. The avatar controller can control the remote avatar, while both the avatar controller and viewer can talk to each other.

Demonstrative pronouns strongly depend on viewpoint. Researchers in the field of psycholinguistics have investigated how people select a viewpoint when using demonstrative pronouns [13]. Ullmer-Ehrich found that a person sets her/his viewpoint at the door when she/he is asked to explain the contents of her/his room [14]. Klein found that a person changed her/his viewpoint along a path when she/he explained the route to a place [15]. The speakers imagined where they were according to the explanation and set their viewpoints on the path. Klein concluded that this change in the viewpoint was the result of a joint viewpoint. Moreover, Imai et al. found that physical manipulation affected the viewpoint of the speaker [5]. A person set her/his viewpoint on a robot when she/he controlled the robot by hand, and communicated with the robot based on a joint viewpoint. This finding means that manipulation causes the emergence of the joint viewpoint. We show the controlling-user creates similar joint viewpoint by virtually controlling remote avatar in this paper.

3 PROT AVATAR

3.1 System Overview

Our prototype system named PROT AVATAR consists of avatar controller and viewer side components (Figure 2). The avatar controller side components consist of a computer, a camera, a microphone, and a speaker (Figure 3). Avatar control software is running on the computer. Its user interface is discussed in this paper. The camera tracks and captures the face of the avatar controller, which is remotely used as an avatar's face (Figure 1). The microphone captures the avatar controller's voice, and the speaker emits the avatar viewer's voice. On the other hand, the avatar viewer side components consist of PROT, a camera, and a microphone. PROT is a video and audio projection system that utilizes a projector with an actuated mirror and an actuated ultrasonic modulation speaker [3]. The ultrasonic modulation speaker can generate a sound source at a distant location [4], and the actuators enable the system to change the location of image

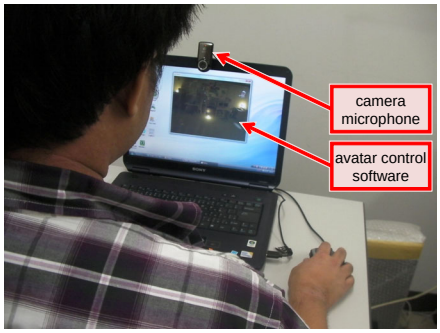


Figure 3: Avatar Controller Side Components

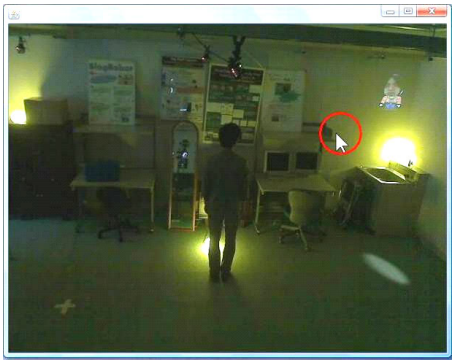


Figure 4: Screenshot of Avatar Control Software

projection and the sound source. The avatar viewer can hear the avatar controller's voice as if it comes from the avatar's location. The camera acquires the scene of the avatar viewer side, which is displayed on the control interface of the avatar controller (Figure 4). The microphone acquires the avatar viewer's voice, which is sent to the avatar controller.

As a basic interaction design, the avatar controller can control the avatar's location by specifying a point in the remote scene (Figure 4), while both the avatar controller and viewer can talk to each other with their microphones and speakers. When the avatar controller completes an operation, actuator control information is sent to the avatar viewer side components, and the avatar- and sound-projection location moves. In the following section, we describe a preliminary study on avatar control policy.

3.2 Preliminary Study

We conducted a preliminary study on avatar projection location. Since PROT employs a projector to project an avatar, the surface of a wall, floor, or ceiling where the avatar is projected can affect the visibility of the avatar. Specifically, we aim to investigate user preference in terms of the background of the projection and distance to the referred points.

Figure 5 shows the study setting. We set up three desktop lights (Figure 6) near the wall for this study. The first one was in front of a white wall. The second

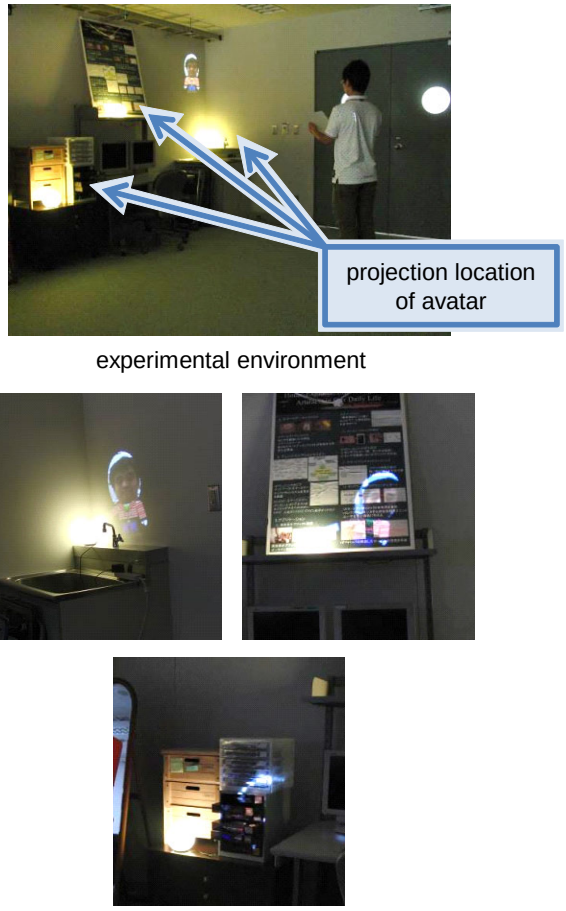


Figure 5: Preliminary Study Setting. The participant saw the projection of PROT AVATAR and determined how much of the avatar the participant could see and how well the participant could identify the location pointed by the avatar.

Figure 6: Desktop Light Used in the Preliminary Study and the Experiment

one was in front of a poster. The third one was in front of a shelf. The white wall corresponds to a flat and solid color background. The poster corresponds to a flat but colorful background. The shelf corresponds to an uneven background.

We invited five volunteer students and showed each of them the avatar as it was moved toward and pointed the target item. After the participant saw the

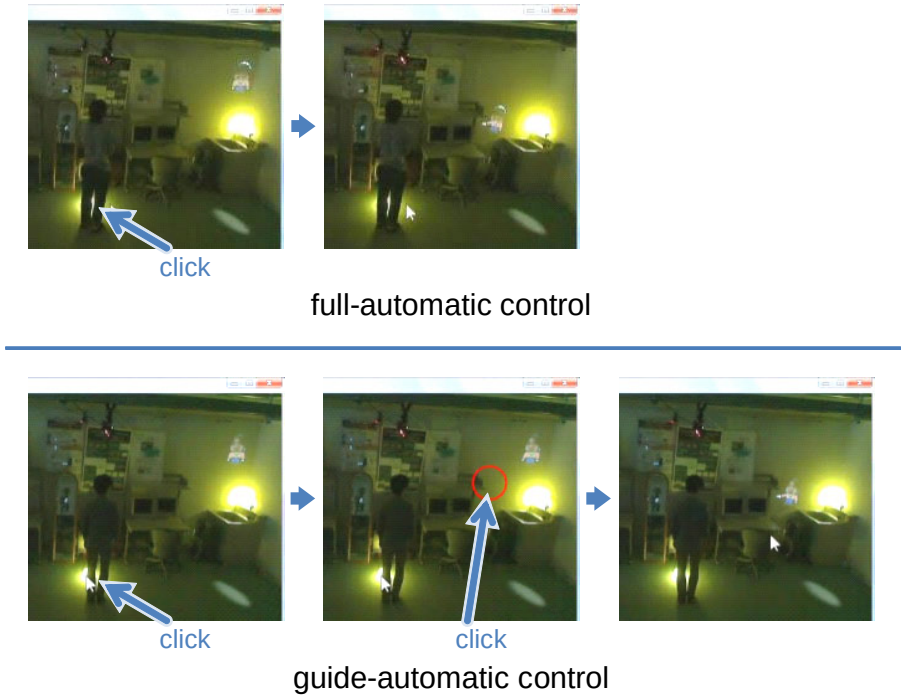


Figure 7: Avatar Control Methods. With full-automatic control, the avatar just moves toward the appropriate location when the avatar controller specifies a pointing location. With guide-automatic control, the appropriate location is first indicated when the avatar controller specifies a pointing location, and the avatar moves after the avatar controller clicks the indicated location.

avatar's movement and pointing gesture for each condition, the participant was asked to score, on a 7-point Likert scale, how well she/he could see the avatar and how well she/he could identify the location pointed by the avatar. We repeated this evaluation procedure for three avatar locations.

For both questions, all five participants scored highest for solid background, followed by colorful and uneven backgrounds. After scoring, the experimenter interviewed the participants and found that the preference for pointing from a solid background was remained even if the avatar's location was far from the target, as opposed to pointing from a location close to the target, on colorful or uneven background.

Thus, here is an avatar control guideline. The appropriate location of the avatar projection is not necessarily near the target location. The system should automatically control the avatar to the appropriate location or recommend the appropriate location, when the avatar controller specifies the target location.

3.3 Design Variations

Based on the result from the preliminary study, we designed full-automatic and guide-automatic control methods (Figure 7). For both methods, the system automatically calculate an appropriate location for avatar projection as follows when the avatar controller specifies the reference target by clicking the control interface. The system first regards flat and solid-color surfaces as candidates of an appropriate location. The system then excludes outside of the movable area caused by actuator limits from candidate areas. The system finally searches the nearest location to the target from candidate areas. Using the calculated appropriate avatar location, the full- and guide-automatic method controls avatar movement.

The full-automatic control method offers an interface that does not require the avatar controller to control where the avatar is projected. When the avatar controller clicks the pointing target on the remote scene, the system automatically moves the avatar to the calculated location, making a pointing gesture.

The guide-automatic control method requires a little more operation by the avatar controller. When the avatar controller clicks the pointing target on the remote scene, the system shows a circle at the calculated location on the screen, prompting the next operation. After the avatar controller clicks in the circle, the system moves the avatar to the location, making a pointing gesture.

4 Experiment

Working with Japanese participants, we compared the full-automatic and guide-automatic control of the avatar in terms of which viewpoint participants take. We can investigate which viewpoint participants take by acquiring the demonstrative pronouns they use,

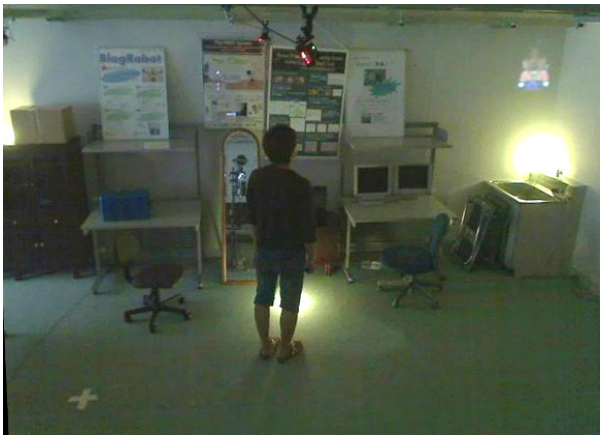


Figure 8: Avatar Viewer Side Environment in the Experiment

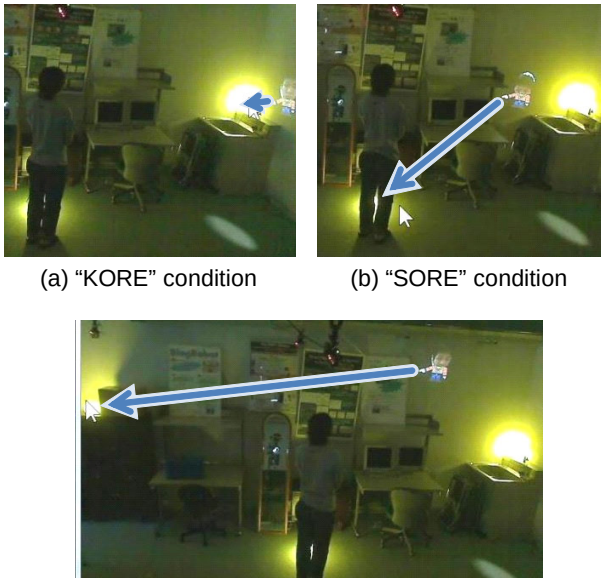
since the use of Japanese demonstrative pronouns is sensitive to the viewpoint of a speaker [5]. For the avatar viewer in remote place, it seems natural that the avatar controller speaks as if her/his viewpoint would be at the avatar's location because the avatar controller's image appears in front of the avatar and the voice of the avatar controller sounds from the avatar's location. Therefore, the avatar viewer expects to interpret the avatar controller's demonstrative pronouns based on the avatar's viewpoint.

4.1 Procedure

Figure 8 shows the remote environment in this experiment. The each participant is first invited to the remote room. In the remote room, the functions of PROT AVATAR are demonstrated during a short conversation between the avatar-viewing participant and the avatar-controlling experimenter. After that, another experimenter invites the participant to the avatar controller side room, and there the experimental session starts.

In the avatar controller side room, the participant is instructed on how to control the avatar using one of the full-automatic or guide-automatic control methods. After that, the participant is asked to arbitrarily try controlling the avatar until the participant sufficiently understands the control method. After the participant reports that the participant understands the control method, the experimenter explains the task for this experiment. There are three desktop lights (Figure 6) in the remote room. The participant is instructed to indicate one of the three lights by moving the avatar and to ask the remote experimenter to change the color of the light using demonstrative pronouns. The remote experimenter will change the color of the light according to the avatar-controlling participant's instruction, using a remote control.

In Japanese, "KORE" is used to refer to an object that is near to a speaker. "SORE" is used to refer an object that is near to a listener. "ARE" is used to refer an object that is far from both a speaker and a listener. Figure 9 shows locations of the avatar that the system determined for each light. From the avatar's viewpoint, it is natural to say "KORE," "SORE," and "ARE" for the case shown in Figures 9(a), 9(b), and 9(c), respectively. Hereafter, we refer to the conditions of Figures 9(a), 9(b), and 9(c) as "KORE" condition, "SORE" condition, and "ARE" condition, respectively.



(a) "KORE" condition (b) "SORE" condition

(c) "ARE" condition

Figure 9: Avatar and Target Location for Each Condition. Blue arrows in the figure indicate which light the avatar is pointing. We set up difference distance from the avatar and the avatar viewer to the target object for each condition.

From the avatar controller's viewpoint, using "KORE" is natural for any location because any location on the screen is close enough to the participants. Therefore, we consider that we can investigate which viewpoint the participants take by analyzing the demonstrative pronouns the participants use.

4.2 Participants and Detailed Settings

36 Japanese students and workers ranging from 19 years old to 43 years old were invited to participate in this experiment. Seven were female and 29 were male. Participants were paid 3000 yen, or roughly 40 US dollars. Following a between-participant experimental design, participants were divided into two groups: full-automatic and guide-automatic control methods. The order of the target lights and color changes were counterbalanced to avoid a learning effect. In this experiment, the appropriate projection area is manually calibrated. The system moves the avatar to or recommends a suitable location in the appropriate projection area based on the specified target location.

Table 1: Experimental Result. The shaded cells indicate the most natural utterances for the remote user.

	"KORE" Condition		"SORE" Condition		"ARE" Condition	
Utterance	Full	Guide	Full	Guide	Full	Guide
"KORE"	18	17	14	4	13	6
"SORE"	0	1	4	11	1	2
"ARE"	0	0	0	3	4	10

Table 2: Counting of Natural Demonstrative Pronouns for the Avatar Viewer. Significantly different ($p < 0.05$) in "SORE" condition and difference trend ($p < 0.1$) in "ARE" condition.

	"KORE" Condition		"SORE" Condition		"ARE" Condition	
Utterance	Full	Guide	Full	Guide	Full	Guide
natural	18	17	4	11	4	10
others	0	1	14	7	14	8
	(n.s.)		($p < 0.05$)		($p < 0.1$)	

4.3 Result

Table 1 shows utterance counts from the experiment. Each row shows the number of participants that used "KORE," "ARE," or "SORE" out of the 18 participants in each condition. The shaded cells indicate the most natural utterances for each condition. Table 2 shows counting of the most natural utterances used by the participants for each condition. In "SORE" and "KORE" conditions, the number of participants that used the most natural utterance for the remote user increased from the full-automatic control to the guide-automatic control. We examined the proportion of participants that used the most natural terms between the full-automatic and guide-automatic control methods, and we found a significant difference ($p < 0.05$) in "SORE" condition and a difference trend ($p < 0.1$) in "ARE" condition, using Fisher's exact test. We found no statistical difference in "KORE" condition. This result indicates that avatar controllers with the guide-automatic control method were more likely to take the avatar's viewpoint more, which appears more natural to the remote user.

5 Discussion and Limitations

The results indicate that fully automatic control of the avatar is not always appropriate for an avatar-mediated telecommunication system. In particular, the experimental result showed the viewpoint of avatar-mediated interaction seemed more natural with guide-automatic control. The avatar controllers having full-automatic control tend to take the users' own viewpoint, which relies on the spatial relations depict on the screen. An interaction design that encourages the user to be aware of the avatar's presence can avoid this problem.

Our findings contribute to the design of the avatar control. The designer of the avatar-based telecommunication system should carefully develop the way of the avatar control not to differ viewpoints between the controlling user and the controlled avatar.

For an avatar-controlling user, the guide-automatic control method requires one more click than full-automatic control method. However, the projection location of avatar is automatically calculated by the system, and the avatar-controlling user does not need to consider where to project. Therefore, compared to manual control, the guide-automatic method requires much less manipulation of the user.

The camera we used in the avatar viewer side place is only a single camera, which did not directly provide the avatar's viewpoint. A multi-camera system installed in the avatar viewer side environment can solve this viewpoint problem in a different way. We think, however, that we have presented a reasonable situation. Our prototype is a single, packaged system, and can be applied in a broader environment.

Although we consider that view point changes similar to those noted by [14, 15] in avatar-mediated interaction may appear for non-Japanese individuals, the methodology we use in this paper is based on Japanese specific rules. Another methodology may be required for speakers of a different language.

6 Conclusion

We conducted an experiment with an avatar-mediated telecommunication system to investigate the viewpoint of an avatar-controlling user. Comparing full-automatic and guide-automatic avatar control, we found that the avatar-controlling participants selected more natural terms for an avatar viewer with guide-automatic avatar control. The results also indicated that an interaction design that encourages the user to be aware of the avatar's presence can help natural communication. We believe our findings contribute in designing an avatar-based communication system.

We plan to implement automatic acquisition of available background by image processing as the next step of system development. We also plan to conduct further experiment to understand human nature in using avatar-mediated telecommunication systems.

References

[1] Yamazaki, K., Yamazaki, A., Kuzuoka, H., Oyama, S., and Miki, H.: Development of an Embodied Space to Support Remote Instruction, *Proceedings of European Conference on Computer Supported Cooperative Work*, pp. 239–258 (1999)

[2] Reitmayr, G., Eade, E., and Drummond, T. W.: Semi-automatic Annotations in Unknown Environments, *Proceedings of IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 1–4 (2007)

[3] Fujimura, R., Nakadai, K., Imai, M., and Ohmura, R.: PROT - An Embodied Agent for Intelligible and User-Friendly Human-Robot Interaction, *Proceedings of International Conference on Intelligent Robots and Systems*, pp. 3860–3867 (2010)

[4] Ishii, K., Yamamoto, Y., Imai, M., and Nakadai, K.: A Navigation System Using Ultrasonic Directional Speaker with Rotating Base, *Proceedings of International Conference on Human-Computer Interaction, Lecture Notes in Computer Science*, Vol. 4558, pp. 526–535 (2007)

[5] Imai, M., Hiraki, K., Miyasato, T., Nakatsu, R., and Anzai, Y.: Interaction With Robots: Physical Constraints on the Interpretation of Demonstrative Pronouns, *International Journal of Human-Computer Interaction*, Vol. 16, No. 2, pp. 367–384 (2003)

[6] Xbox avatars: http://www.xbox.com/en-US/live/avatars/

[7] Second Life: http://secondlife.com

[8] Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K., and Mitsuishi, M.: GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions, *Proceedings of ACM Conference on Computer Supported Cooperative Work*, pp. 155–162 (2000)

[9] Ishiguro, H., and Nishio, S.: Building artificial humans to understand humans, *Journal of Artificial Organs*, Vol. 10, No. 3, pp. 133–142 (2007)

[10] Shiomi, M., Sakamoto, D., Kanda, T., Ishi, C.T., Ishiguro, H., and Hagita, N.: A Semi-autonomous Communication Robot: A Field Trial at a Train Station, *Proceedings of ACM/IEEE International Conference on Human-Robot Interaction*, pp. 303–310 (2008)

[11] Nakanishi, H., Murakami, Y., Kato, K.: Movable Cameras Enhance Social Telepresence in Media Spaces, *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 433–442 (2009)

[12] Nakanishi, H., Kato, K., Ishiguro, H.: Zoom Cameras and Movable Displays Enhance Social Telepresence, *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 63–72 (2011)

[13] McNeill, D.: Psycholinguistics: A New Approach, Harper & Row Press (1987)

[14] Ullmer-Ehrich, V.: The structure of linving space descriptions, *Speech, Place and Action*, pp. 219–250 (1982)

[15] Klein, W.: Local Deixis in Route Directions, *Speech, Place and Action*, pp. 161–182 (1982)