Preliminary Evaluation of a Telepresence Robot Conveying Pre-motions for Avoiding Speech Collisions

Komei Hasegawa[1] Yasushi Nakauchi[1]

[1] University of Tsukuba

Abstract: In this paper, we propose a telepresence robot for avoiding speech collisions which occur in remote conversations. In face-to-face conversations, humans predict who will speak in next by referring other participants ' motions. These motions are called pre-motions. However, pre-motions are likely to be unperceived in teleconferences such as video chat conversations using two-dimensional video image. The failure of the prediction causes speech collisions. It undermines the participants' motivation and wastes time. To solve this problem, we developed a telepresence robot that conveys gestures including pre-motions. We use Kinect for the gesture perception device and humanoid robot for conveying unconscious gestures. According to a preliminary experiment we conducted, our proposed method has effect to reduce speech collisions.

1 Introduction

In daily life and business situation, video chat systems are widely used to communicate with remote participants. However, in such meetings, the speech collisions (two or more participants start speaking at the same time), happens very often. According to the report by Tamaki et al., the speech collisions at web meeting occur about 30 times as many as that of face-to-face conversations [1]. The speech collisions undermine the participants' motivation and waste time.

On the other hand, humans can conduct smooth turn-taking in face-to-face situations. It is because humans are able to exchange non-verbal cues. Especially in turn-taking, non-verbal information is used as a cue, which enables participants to predict the next speaker. This cue is called pre-motion. According to Marjorie, instances of pre-motion are leaning forward, unfolding arms, turning the body over, and so on [2]. It is assumed that humans express these behaviors unconsciously. Furthermore, it is supposed that listeners receive some information from not only speaker ' s gestures expressed by consciously but also unconscious gestures [3]. Thus, it is important to convey unconscious gestures for avoiding speech collisions.

Though, web meetings are used in many occasions, it is difficult to exchange non-verbal cues including pre-motions in web meetings since two-dimensional vision could convey poor presences. Therefore, speech collisions tend to be appeared in conventional teleconferences. On the other hand, while it is said that embodied robots make people feel stronger presence than visions [4]. Therefore we think robots could be used for conveying the presences to remote sites.

In this paper, we propose a telepresence robot for reducing speech collisions shown in Figure 1. We use motion capture system as a method to operate the robot's motion. This enables to input unconscious gestures and convey the pre-motions. We expect that these pre-motions expressed by the robot help us to reduce speech collisions.

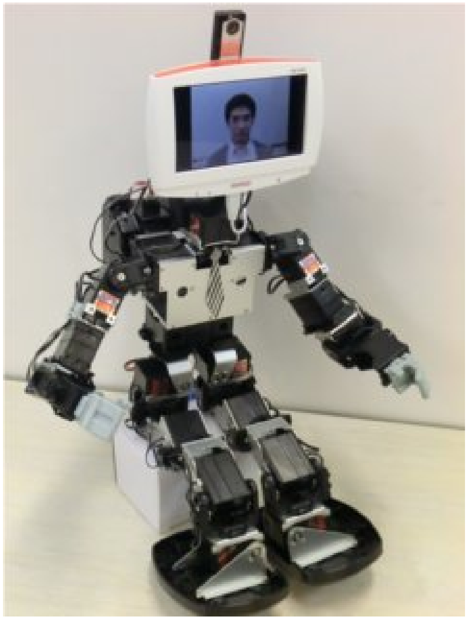

Figure 1: Telepresence robot

2 Related Works

Some existing telepresence robots for remote communications were developed. Paulos et al. developed telepresence robot named PRoP [5]. It had a camera, microphone, speaker, and display on mobile robot. It allowed teleoperator to communicate and roam around. This existing research reported that the robot provided social embodiments. Similar telepresence robots, QB by Anybots and RP-7 by InTouch Technologies, were already commercialized

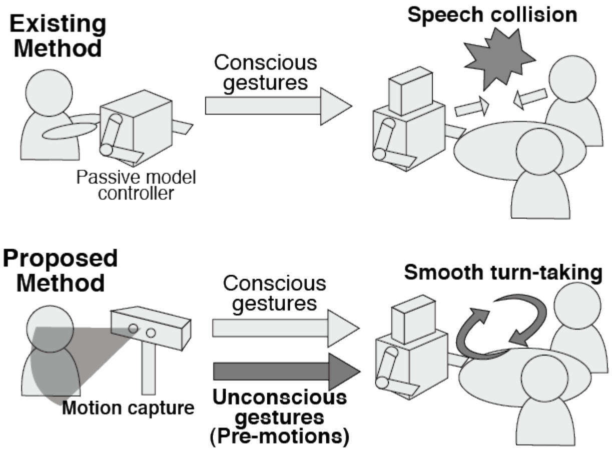

Figure 2: Concept of our method

Figure 3: Kinect, a motion capture device

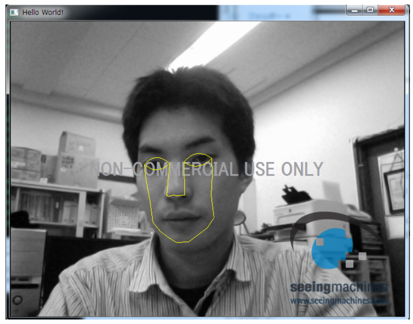

Figure 4: A view of faceAPI interface

and in practical use [6 7]. However, these telepresence robots had no movable neck and arms. Therefore, it was impossible to convey expressive head and hand gestures.

On the other hand, MeBot having a movable neck and arms enabled communications using various kinds of head and hand gestures [8]. However, the operation method for robot ' s arms was that the operator manipulated the passive model controller. This method was unable to express the operator's unconscious gestures.

3 System Design

3.1 Concept

Figure 2 shows the concept of our proposed method. In existing method, a remote participant operated a telepresence robot by changing a posture of a passive model controller. It is assumed that this method can perceive only the gestures, which the remote participant tried to express consciously. Therefore, the conscious gestures, which conveyed via the robot has no pre-motions. Then, the participants are unable to avoid speech collisions in the conference.

On the other hand, our proposed method employs motion capture to perceive the remote participant's conscious and unconscious gestures. Then, unconscious gestures including pre-motions are conveyed. It is expected that the conveyed unconscious gestures help the participants to avoid speech collisions and make the smooth turn-taking.

3.2 Telepresence Robot

Figure 1 shows the view of proposed telepresence robot. We designed the robot to convey non-verbal cues such as arm motions, head motions, looking directions, body directions, and facial expressions. We employed KHR-3HV that is a desktop sized humanoid robot by Kondo Kagaku Company as the robot ' s platform. It was built with extra servomotors for our research.

The robot had two 4 DOF arms which enables various kinds of hand gestures. It also had a 2 DOF waist, which allows leaning forwarding motions and changing body directions. For expressing head motions and looking directions, it had a 3 DOF neck. As the robot ' s face, a small size display was fixed on the neck. The display showed the operator ' s face video so that it conveyed the facial expressions. Using its arms, waist, neck and face, the robot expressed operator ' s conscious and unconscious gestures.

3.3 Control Interface

3.3.1 Motion Capture Controller

For controlling the robot's arms and waist, we utilized Kinect by Microsoft Corporation seen in Figure 3. By combining Kinect for Windows SDK, it enabled to capture the operator ' s motion without any markers. With using the motion capture such as Kinect, this system enable to input the gestures whether conscious or unconscious.

For controlling the robot ' s head, we employed a web camera and FaceAPI by Seeing Machines seen in Figure 4. FaceAPI is a software library for recognizing a face direction in real time.

3.4 System Configuration

Proposed system's configuration is seen in Figure 5. A remote participant attended the meeting via the telepresence robot. The remote participant ' s motion such as arms, waist, and head, which sensed by the methods previously described were sent to the computer at

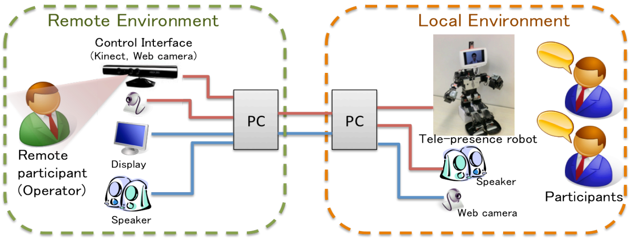

Figure 5: System configuration

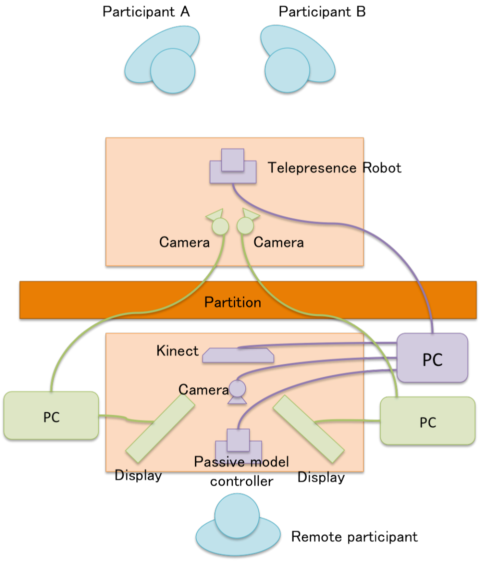

Figure 6: Experimental environment

the local place. The motion data were translated and expressed by the robot with the facial video. The participants’ appearances and voices were captured by a web camera and a microphone. Then, those were expressed to the remote participant by the display and the speaker.

4 Preliminary Experiment

We conducted a preliminary experiment to confirm that our telepresence robot allows to reduce speech collisions. We set up teleconferences. An overview of the experimental environment can be seen in Figure 6. A group of three participants attended the teleconference. One participant of the group seated in a partitioned place. He/she was allotted as a remote participant and joined the conference via the telepresence robot.

In each session, they had four minutes long conversations. We assigned the participants a talking task in order to bring about active conversations. We used *The Desert Survival Problem* which we modified for this experiment.

Figure 7: Passive model controller

4.1 Implementation for Experiment

To compare our gesture perception method with existing one, we implemented a passive model controller. The view of implemented passive model controller is shown in Figure 7. The controller was a passive model of the telepresence robot and had same DOF. An operator could change the posture of the passive model using his/her hands, and then the telepresence robot also changed the posture. This control method was inspired by MeBot[8].

4.2 Three Modes for Comparison

We implemented the experiment in following three Modes.

Mode A: Motion capture controller mode

Mode B: Passive model controller mode

Mode C: Face-to-face mode

In Mode A which is our method, the remote participant operated the robot with the motion capture. Similarly in Mode B, the robot was operated with the passive model controller. In Mode C, all of three participants attended the face-to-face conference. Therefore, no one was allotted as a remote in Mode C. They performed two sessions, Modes A and B, for each of remote participant. In addition this, one face-to-face session was performed. Therefore, each group had totally 7 sessions.

4.3 Experimental Result

A total of three groups (9 people) participated the experiment.

We conducted subjective evaluations of the sessions. We asked remote participants to score three evaluation items after each session. In the case of Mode C, all participants were asked to evaluate the items. The evaluation was done by a five-point scale ranging from 1(disagree) to 5(agree). They evaluation items were ”I feel it was easy to join the conversation,”

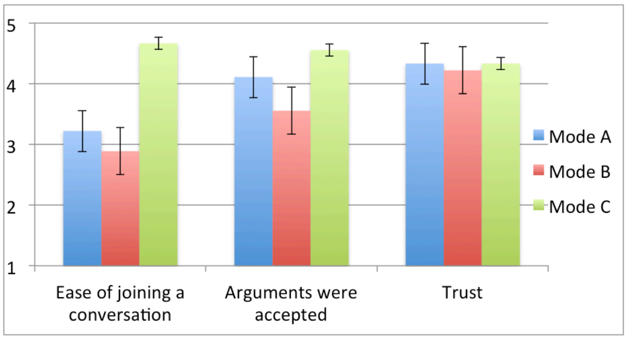

Figure 8: Average ratings for subjective evaluations

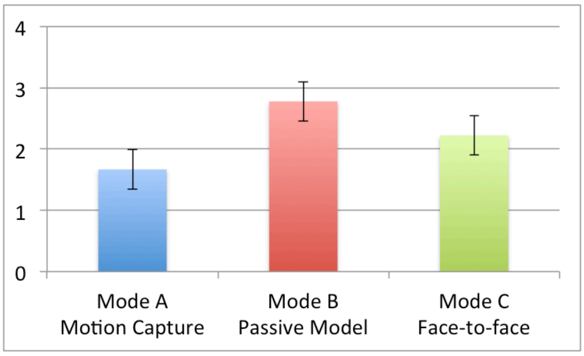

Figure 9: Average counts of speech collisions

"I feel my arguments were accepted," and "I trusted other participants". Figure 8 shows the averages and standard deviations.

Further more, we recorded videos of all sessions with two video cameras, for local and remote site. By reviewing the videos, we counted speech collisions in each session. Figure 9 shows the average counts of speech collision occurred in each Mode.

4.4 Discussion

According to the results of subjective evaluations, "ease of joining a conversation" and "arguments were accepted" got positive evaluations in Mode C. These results indicated that face-to-face communications were easier to joining and arguing than remote. It is assumed that remote participants tend to be failed to notice. Discussing the remote conversations, Mode A was evaluated more positive than Mode B in these two items. This shows that our method make remote participants feel easier to communicate. It is supposed that unconscious gestures conveyed by the proposed telepresence robot attracted other participant.

The average counts of speech collisions in Mode A were less than Mode B. It is assumed that unconscious gestures and pre-motions expressed by our method help participants to avoid speech collisions.

Since the experiment was preliminary one, participants were only three groups (9 people). To evaluate the method more accurately, we need to conduct the experiment with more subjects.

5 Conclusion

In this paper, we proposed the telepresence robot conveying unconscious gestures for avoiding speech collisions. To convey unconscious gestures, we employed motion capture system for robot's gesture control interface. We implemented this method by using Kinect.

In order to confirm the advantages of our proposed method, we conducted a preliminary experiment. We arranged three Modes, Mode A which was our method using motion capture controller, Mode B which was existing method using passive model controller and Mode C which was a face-to-face situation. The experimental results shows that our method reduced speech collisions and made remote participants feel easier to joining conversations than existing one. Therefore, it is assumed that conveying unconscious gestures by the telepresence robot has effect to avoid speech collisions and to enable smoother telecommunications. To evaluate the proposed method more accurately, we will conduct the experiment with more groups.

References

[1] Hidekazu Tamaki, Suguru Higashino, Minoru Kobayashi, Masayuki Ihara: Reducing Speech Contention in Web Conferences, *2011 IEEE/IPSJ International Symposium on Applications and the Internet*, pp.75-81 (2011)

[2] Marjorie F. Vargas: Louder Than Words: An Introduction to Nonverbal Communication, *Iowa State Press* (1986)

[3] J. Cassell, D. McNeill, K. E. McCullough: Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information, *Pragmatics and Cognition*, Vol.7, No. 1, pp. 1-33 (1999)

[4] J. Hauber, H. Regenbrecht, A. Hills, A. Cockburn, and M. Billinghurst: Social presence in two- and three- dimensional videoconferencing, *The 8th Annual International Workshop on Presence* (2005)

[5] E. Paulos, J. Canny: Social Tele-embodiment: Understanding Presence, *Autonomous Robots*, Vol. 11, No. 1, pp. 87-95 (2001)

[6] Anybots: QB, https://www.anybots.com

[7] InTouch Technologies: RP-7, http://www.intouchhealth.com

[8] S. O. Adalgeirsson, C. Breazeal: MeBot: A Robotic Platform for Socially Embodied Telepresence, *HRI2010*, pp. 15-22 (2010)