

Using the iCat as Avatar in Remote Meetings

Dirk Heylen, Mannes Poel, and Anton Nijholt

Human Media Interaction, University of Twente
PO Box 217, 7500 AE Enschede
The Netherlands
{heylen,mpoel,anijholt}@ewi.utwente.nl

Abstract. We compared two ways of remote participation in a meeting. One in which a video-connection existed between the remote participant and the collocated participants and one in which the remote participant was represented by an iCat. We asked the participants to rate the conversations on various dimensions. The remote participants tended to prefer the meetings with the iCat whereas the co-located participants preferred the video connection on most dimensions.

1 Introduction

In this paper we present an analysis of recordings of three-party conversations in which two participants were co-located in the same room and the third participant was in another room. We made recordings of two conditions, which we will refer to as VIDEO and ICAT. In the first condition, we used two cameras in the room of the co-located participants, each directed at one of the participants (see Figure 1). The two video streams were displayed on a computer screen in front of the remote participant. In this condition, the co-located participants could see the image of the remote participants on a computer screen. The image was taken from a webcam placed on top of the computer screen of the remote participant. This set-up more or less resembles a typical video-conferencing situation.

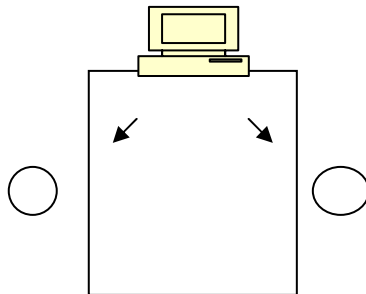


Fig. 1. Schematic view of the co-located participants' room in the VIDEO condition. Circles represent co-located participants sitting at opposite sides of the table (the rectangle), with two cameras pointed at them (the arrows). In the ICAT condition the computer screen is replaced by an iCat and the cameras are removed.

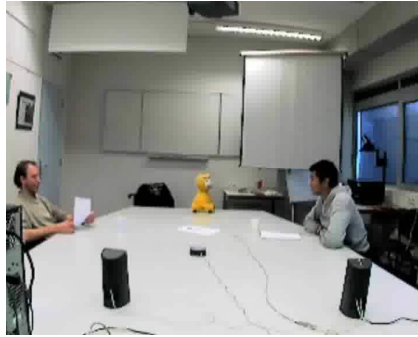


Fig. 2. The ICAT condition

In the ICAT condition, the computer screen in the co-located participants' room was replaced by an iCat¹ that was controlled by the remote participant. The iCat thus functioned as an avatar representing the remote participant. For the remote participant the view on the co-located participants' room now came from the camera positioned in the nose of the iCat.

The remote-participant had several options available to control the iCat. The head of the iCat could be turned in the direction of either of the participants or to the middle of the room. The participant could also have the iCat shake or nod its head and could choose to make the iCat look amazed (raised eyebrows) or smile.

Comparing settings of remote communication in this way allows us to study the effects of manipulating technologies that mediate the communication process comparing the different affordances that each of them offers. The effects we are interested in concern changes in how the conversation proceeds and changes in the appreciation of the communication by the participants. What do participants do differently? What parts of the technological affordances do they use? What do they like or dislike about the opportunities? Our goal is to reach a better understanding of face-to-face conversation and of the need for facial expressions, gaze in processes such as grounding and addressing. This understanding may allow us to improve upon the ways to communicate remotely.

The paper is structured as follows. In the next section, we look at the two conditions in more detail in terms of the affordances offered and present some conjectures about how we expect the participants to behave and appreciate the settings. Next, we present details about the experiment and the measures. This is followed by a section on results and analysis.

2 Remote Presence

Mediated communication far outnumbers face to face conversations in our present lives. It can take the form of exchanges of information through telephone, e-mail, newspapers, television, chat, podcasts, blogs, tele- and videoconferencing, or involve meetings in 3D worlds such as Second Life or in on-line games. It is a common

¹ http://www.research.philips.com/technologies/syst_softw/robotics/index.html

observation that these mediated forms of communication offer less rich interactions than face-to-face interactions, mainly because of the inability to convey nuances through nonverbal means of communication and gaze, or because the intimate feedback loop between interlocutors is tampered with in these forms of communication ([2,4,6,7,8]. In studies on Computer Mediated Communication [9] these differences are typically seen as shortcomings and new technologies are introduced to overcome them.

The differences between face-to-face and mediated conversation can be described in terms of the 10 dimensions that characterize direct spontaneous conversation according to [1]: co-presence, visibility, audibility, instantaneity, evanescence, recordlessness, simultaneity, extemporaneity, self-determination and self-expression. In the two conditions that we are studying, one of the participants is remote and thus differs with respect to face-to-face conversations on the dimension of co-presence. In the VIDEO condition, the remote participant can see the others and vice versa through a video connection. However, the view is fixed through the camera-position and the view of the environment is thus severely limited. Eye-contact is also not possible. In the ICAT condition, the co-located participants cannot see the remote participant, but can only see the iCat perform the instructions that are given to it by the remote participant. The view of the remote participant is limited to three options: looking at one participant, the middle of the table or to the other participant. In contrast to the VIDEO condition, the remote participant can thus only look at one co-located participant at the same time [4,8]. On many dimensions one could argue that there are no special differences with respect to face-to-face interactions, but with respect to the two final ones some peculiar aspects can be noted. The iCat can be thought of as a puppet through which the remote participant speaks to the others. As a puppet, the iCat speaks not for itself (as one might consider in the case of an autonomously operating robot) but as a messenger controlled by the remote participant.

The possibilities of nonverbal expressivity of the remote participant are severely limited to the few facial expressions and head movements that can be made. By being able to turn the head of the iCat in the direction of one or the other co-located participant, the remote participant could signal attention or who is being addressed. Agreement and disagreement or affirming or negating could be signaled by nods and shakes. Facial expressions are limited to attitudinal reactions of enjoyment and surprise. One can expect the two conditions to differ in a number of respects, both in the way the communication proceeds and in the way it is perceived and evaluated. In our current analysis of the data we looked at effects on grounding, addressing, turn-taking and feelings of co-presence [10].

With respect to grounding, the question is whether the participants feel that it is easy to acknowledge contributions and whether it is more difficult to receive acknowledgements of contributions. In the ICAT condition the acknowledgements by the remote participant can be given verbally through nodding and shaking the head of the iCat. Also the facial expressions can provide feedback. All of these, however, need to be consciously executed by the remote participant, whereas in the VIDEO condition, the facial expressions can provide a running commentary to what is being discussed.

Gaze plays an important role in addressing. Goodwin [5] observed how at the beginning of a turn, speakers may try to achieve mutual gaze with the interlocutor to make clear who they address and get acknowledgment of this. In the VIDEO condition

it is not possible for the remote participant to indicate who is the intended addressee by nonverbal means but in the ICAT condition, manipulating the head orientation of the iCat can be used in this way.

The controls for the iCat that were implemented in this set-up do not allow the remote participant to signal turn-requests with the nonverbal means that people usually use in conversations [3]. In the VIDEO condition, posture shifts and other visual cues for turn-taking, are visible to the co-located participants, however.

The effects on grounding, addressing, turn-taking, and other aspects of the conversation contribute to the general perception of co-presence: the feeling of the interactants that they perceive the others and the others perceive them. The iCat and the video connection differ as media in how they are able to connect people. This feeling will often be different for the co-located and the remote participant, particularly if the affordances of the media are not the same for both parties.

3 Data

We invited four teams of three participants to each conduct two meetings². Two of the teams first conducted a meeting in the VIDEO condition followed by a meeting in the ICAT condition. The order was reversed for the other two teams. Each meeting lasted between 10 to 15 minutes. The teams were given separate tasks for the two meetings. The first task concerned the discussion about the preferred buttons to go on a remote control. Each participant was told to defend a particular favorite. The second task involved a discussion about the shape and colour of the control. The remote participant got instructions on how to use the interface to control the iCat and was shown the effects of the actions.

In the next section we discuss the results of our analysis of the data. We will discuss the use of the controls of the iCat by the remote participant and the various questionnaires that were presented to the participants.

4 Results and Analysis

Table 1 shows the number of times a participant (A-D) used a particular control for the iCat. It appears from the table that of the four participants, B deviated considerably from the others in that the others used the controls more than two (and D even three) times as often.

All participants used the head orientation controls the most. For each of them this constituted more than half of the actions. In total 106 out of the total of 177 actions were head movement actions. Left and Right movements were widely preferred above the movement to the center. With respect to the other two head movements, nods and shakes, the distribution is about equal, except for participant C who lets the iCat nod its head 8 times but lets it shake its head only once. The expressive controls for smile and eye-brow raise are used about as often as those for nod and shake, with about the same amount of smiles and eye-brows occurring for each of the participants.

² The experiment was carried out by two BSc students Frans van der Veecken and Feitze de Vries (<http://referaat.cs.utwente.nl/new/papers.php?confid=9>).

Table 1. Each row shows the number of times a particular remote participant used the control for the iCat

<i>Participant</i>	<i>Left</i>	<i>Center</i>	<i>Right</i>	<i>Nod</i>	<i>Shake</i>	<i>Smile</i>	<i>Amazed</i>	Total
A	14	3	14	4	4	6	5	50
B	6	1	4	3	1	3	2	20
C	14	7	11	8	1	3	3	47
D	14	4	14	7	8	6	7	60
Total	48	15	43	22	14	18	17	177

Table 2. Ratings of the quality of the interactions

		ICAT	VIDEO
Influence	C	3.3	3.9
(rate your influence in the meeting)	R	3.8	3.3
Satisfaction	C	2.7	3.2
	R	3.3	2.9
Turn-taking	C	3.4	4.0
(how well did turn-taking go)	R	2.8	3.3
Addressing	C	3.4	4.1
(was it easy to address someone)	R	3.5	2.3
Getting Addressed	C	3.4	4.3
(was it clear you were addressed)	R	3.5	3.3
Involved in conversation	C	3.1	4.4
(rate your involvement)	R	3.5	4.0
Involved in task	C	2.6	4.0
	R	3.5	3.0
Grounding	C	6.3	4.8
(awareness of reactions of others)	R	3.8	4.3
Similar F2F	C	5.4	4.0
	R	4.3	6.5

Our analysis of one of the ICAT meetings showed that the remote participant used the head orientation to direct the camera to the participant that was speaking. One participant was looked at 78% of the time he was speaking and the other was looked at 97% of the time he was speaking.

When asked which controls the remote participants found most useful, they answered *nod*, followed by *shake*, *smile* and *surprise*. They found the interface to control the iCat easy to use and rated it on average 4.25 out of 5.

We asked the participants to fill out questionnaires with various questions. The results are presented in the table below. Participants rated the quality of the meetings in both conditions on a five point scale, except for the last two questions which were rated on a 7 point scale.

The table shows the average ratings for each dimension of the co-located (C) and remote (R) participants separately. For each line the highest figure appears in bold-face. From this table it becomes clear that for all questions except the last two, the

co-located participants prefer the VIDEO condition. The remote participant, on the other hand prefers the ICAT condition, except when asked about turn-taking, involvement in conversation and for the last two questions on awareness of others and similarity to face-to-face. This can be explained by the fact, that the remote participant has no global view of the co-located participants in the ICAT condition (negative effect on awareness) and needs to communicate through an artificial interface in this condition as well (negative effect on similarity). The co-located participants find talking to an anthropomorphic iCat closer to face-to-face conversations than interacting with the remote participant through a video connection.

5 Conclusion and Future Work

In on-line games and virtual communities, people use avatars as representations of themselves to communicate with each other. There have been a few suggestions in the literature to use avatars in remote meetings as well, instead of using the ordinary means of video conferencing. In our studies, we are looking at the effects of manipulating the communicative expressivity of avatars on the appreciation of the communication and on how the conversations proceed differently. In the case of the iCat, we found that the participants controlling the iCat like it better than the participants that are faced with it. The results reported on in this paper were primarily based on the analysis of the questionnaires that the participants filled out. Currently, we are processing the video's to see in more detail how the conversations proceeded, looking at the timing and placement of the nonverbal controls that were offered.

Acknowledgements

This work is supported by the European IST Programme Project FP6-0033812 (AMIDA). This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

References

1. Clark, H.H.: *Using Language*. Cambridge University Press, Cambridge (1996)
2. Clayes, E.L., Anderson, A.H.: Real faces and robot faces: The effects of representation on computer mediated communication. *International Journal of Human-Computer Studies* 65(6), 480–496 (2007)
3. Duncan, S.: Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology* 23, 283–292 (1972)
4. Garau, M., Slater, M., Bee, S., Sassa, M.A.: The Impact of eye-gaze on communication using avatars. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 309–316 (2001)
5. Goodwin, C.: *Conversational organisation, Interaction between Speakers and Hearers*. Academic Press, New York (1981)
6. Isaacs, E.A., Tang, J.C.: What video can and can't do for collaborations: a case study. In: *Proceedings of the first ACM International Conference on Multimedia*, pp. 199–206 (1993)

7. Nowak, K., Biocca, F.: The effect of the Agency and Anthropomorphism on Users' sense of Telepresence, Copresence and Social Presence in Virtual Environments. *Presence* 12(5) (2003)
8. Sellen, A.J.: Remote Conversations: the effects of mediating talk with technology. *Human-Computer Interaction* 10, 401–444 (1995)
9. Whittaker, S.: Theories and Methods in Mediated Communication. In: Graesser, A., Gernsbacher, M., Goldman, S. (eds.) *The Handbook of Discourse Processes*, pp. 243–286. Erlbaum, NJ (2002); Richardson, D.C., Dale, R.: Grounding dialogue: eye movements reveal the coordination of attention during conversation and the effects of common ground, *COGSCY 2006* (2006)
10. Zhao, S.: Towards a Taxonomy of Copresence. *Presence* 12(5), 445–455 (2003)