Conveying Directional Gaze Cues to Support Remote Participation in Hybrid Meetings

Betsy van Dijk , Job Zwiers, Rieks op den Akker, Olga Kulyk, Hendri Hondorp, Dennis Hofs, and Anton Nijholt

Human Media Interaction, University of Twente,
P.O. Box 217, 7500 AE Enschede,
The Netherlands
{bvdijk,zwiers,infrieks,o.kulyk,g.h.w.hondorp,
hofs,anijholt}@ewi.utwente.nl

Abstract. We study videoconferencing for meetings with some co-located participants and one remote participant. A standard Skype-like interface for the remote participant is compared to a more immersive 3D interface that conveys gaze directions in a natural way. Experimental results show the 3D interface is promising: all significant differences are in favor of 3D and according to the participants the 3D interface clearly supports selective gaze and selective listening. We found some significant differences in perceived quality of cooperation and organization, and on the opinions about other group members. No significant differences were found for perceived social presence of the remote participants, but we did measure differences in social presence for co-located participants. Measured gaze frequency and duration nor perceived turn-taking behavior did differ significantly.

Keywords: Hybrid meetings, videoconferencing, selective gaze, selective listening, social presence, group process, turn-taking, remote participation.

1 Introduction

Collaboration between physically dispersed teams has become very important in the last decades, in industry and in science. Because of this development much research has been done on systems for multiparty videoconferencing and many systems are on the market now. Videoconferencing systems such as Skype or Adobe Connect offer a 2D picture-in-picture interface on a single video screen. All the participants are seated facing the camera and are visible in separate video frames. These frames are combined at a central location and the output is broadcast to the participants. Added value of the use of such systems, as compared to phone conferences, is that both speech and facial expressions are communicated. However, due to this setup, such videoconferencing systems fail to support selective gaze and selective listening [13]. Participants cannot show in a natural way to whom they look and they are not aware of who is visually attending to them. Though these disadvantages of distributed meetings with only mediated communication are well-known (e.g., [13]), remote meetings are often used to avoid having to choose between traveling too much or meeting too little.

A. Esposito et al. (Eds.): COST 2102 Int. Training School 2010, LNCS 6456, pp. 412–428, 2011.

In this paper we focus on hybrid meetings, where one remote person is connected to a meeting taking place in a meeting room. Hybrid meetings are interesting because both face-to-face interaction and mediated interaction occur in the same group [2]. Remote participants might, as a consequence of their isolation, feel different about the group, the process and the outcomes of the meeting [2]. This feeling will be strengthened if co-located meeting participants use the opportunity they have to form a cohesive subgroup, making the remote participant a marginal member of the group [2]. As Yankelovich put very aptly in [20]: *"If you have ever dialed-in to a meeting taking place in a conference room, you probably know what it feels like to be a second-class citizen"*. Hybrid meetings suffer from almost all problems of fully distributed meetings but the difference in user experience between the remote participants and the co-located participants results in many additional problems that mainly have to do with social presence [20], the feeling of being together with another. A few important problems of the remote participant are the inability to participate in informal conversations (important for forming relationships and trust) and difficulty to break into a conversation. The people in the meeting room tend to forget about the remote participant because the physical presence of the people in the room takes their attention [20]. These and other problems make it difficult for the remote participant to stay engaged and keep paying adequate attention.

In this paper, we examine the effects of two different user environments for videoconferencing. The environments aim to improve the user experience of a remote participant. We compare a "*standard*" *conventional video conferencing interface* with an interface where video streams were presented to remote participants in an *integrated 3D environment*. In the conventional interface, co-located meeting participants all look straight into their webcam and they are visible to the remote participant in separate video frames presented in a horizontal row and in random order on a classical large screen. In the meeting room the video image of the remote participant is projected on a large projection display at the head of the table. Such a multimodal interface already communicates both speech and facial expressions. However, non-verbal behavioral cues like gaze direction and selective listening are lacking. The integrated 3D interface aims to enhance the group process and social presence of the participants by conveying gaze directions of both co-located participants and the remote participant in a natural way. To accomplish this other camera positions are chosen (explained in Section 3) and the video images of the co-located participants are presented to the remote participant in a more immersive way: they appear to be sitting around a virtual table, in a location that is consistent with the real, physical, situation. The co-located participants are presented to the remote participant on the same classical screen that was used for the conventional interface.

Other research projects have focused on improvement of audio- or videoconferencing systems before, and often conveying gaze direction was an important part of the efforts (e.g., [14, 16, 20]). In contrast to these projects, where they built special systems that often were expensive, we used rather basic low cost equipment (cameras, normal computers, microphones, standard screens). Hence the environments are easy to realize and change once the software is available.

This research takes place in the context of the European Network of Excellence on Social Signal Processing (SSPNet) and the European Augmented Reality Multi-party Interaction project (AMI and its successor AMIDA). SSPNet focuses on recognition,

interpretation and synthesis of non-verbal behavioral cues in data captured with sensors like microphones and cameras. The aim is to provide computers with the ability to sense and understand human social signals and to design computer systems capable of adapting and responding to these signals. AMIDA concentrates on multi-party interaction during meetings and aims to develop technologies that can provide live meeting support to remote and co-located meeting participants. Part of the work is capturing non-verbal meeting interactions (posture, gestures, head orientation) and to look at ways to transform these into a virtual reality representation of a meeting room and meeting participants [9]. The real-time display of, for instance, head orientations will not always display gaze direction accurately, but it allows a fairly realistic representation of the focus of attention of participants (e.g., looking at a speaker, addressing someone). Within AMIDA we developed a demonstrator system to support remote meeting participation [1]. This User Engagement and Floor Control (UEFC) demo uses, amongst others, automatic speech recognition, visual focus of attention recognition and addressee detection. It can automatically support (remote) participants in identifying (1) if they are being addressed and (2) who is speaking to whom. The graphical user interface of the UEFC demo presents an overview of the meeting room and separate video images of the faces of the other participants. Although the design of the interface was not the focus of [1], an important observation was that participants appreciated the overview and the separate images of faces but had difficulty establishing mutual gaze in remote interactions. The experimental user environments that are evaluated in this paper were inspired by the experiences with the UEFC demo and aim to improve mutual gaze in interactions.

This paper presents the effects of the two experimental environments on perceived social presence and perceived quality of the group process and satisfaction with the outcome. In addition, subjective data will be presented on the turn-taking process, recognition of gaze behavior (awareness of who is looking to whom) and usability of the environments, as well as preliminary results of the analysis of objective data on gaze behavior. The paper is organized as follows. Section 2 presents related work. In Section 3 we describe the design of the two experimental conditions in more detail and we present the hypotheses. Section 4 describes the methodology used in the user study we conducted and Section 5 gives the results, followed by a discussion of the results in Section 6. Finally, conclusions and future work can be found in Section 7.

2 Transmission of Gaze Behavior in Mediated Communication

Studies comparing mediated communication with face to face communication often point out the importance of non-verbal behavior (facial expressions, head nods, gaze and gestures) for turn-taking and for the transmission of social and affective information [17]. To be conveyed, non-verbal behavior depends on the presence of visual information. Hence it is to be expected that technologies that do not support visual information show impaired communication [19]. However, simply adding a video channel to the supporting technologies does not always result in improved mediated communication that resembles face to face communication more in the sense that it is more efficient. Whittaker [18, 19] argues we should identify the contributions of various communication behaviors (e.g., speech, gaze, gestures, backchannel feedback) in

supporting core communication phenomena, such as turn-taking. If we can also specify how the affordances of various mediated communication technologies affect behaviors we will be able to predict more precisely how technologies affect communication [19]. Much research has been done to elucidate the role of gaze in mediating turn-taking behavior in face to face communication (e.g., [7]). One hypothesis that Whittaker [19] treats is that technologies that do not transmit gaze behavior properly will disrupt turn-taking.

Results of studies that investigated this hypothesis are mixed. Sellen [13] found no differences between an audio-only system and three different videoconferencing systems on measures of turn-taking behavior, such as duration of turns, turn frequencies, number of interruptions. Compared to face to face communication, both the audio-only system and all videoconferencing systems showed reduced ability of listeners to take the floor spontaneously (less interruptions) and speakers used more formal techniques to hand over the floor (e.g., naming a possible next speaker). However, subjective data gathered by questionnaires did show differences in perceived influence of the systems. Participants mentioned several benefits of video. Video was thought to: (a) lead to more natural and more interactive conversations; (b) help identify and discriminate among speakers and to help to generally keep track of the conversation; (c) allow one to determine whether others are paying attention; (d) may support selective gaze and selective listening; (e) make them feel more part of the group and less remote from the other participants.

Vertegaal [16] also argues that just adding video to increase the number of cues conveyed does not necessarily improve communication when it comes to regulation of conversations. He believes turn-taking problems with multiparty conferencing systems may be attributed to a lack of cues about other participants' attention. He developed a system, the GAZE Groupware System, which provides awareness about the participants' gaze direction. By conveying only gaze direction the system allowed meeting participants to establish who is talking or listening to whom without some of the drawbacks of videoconferencing systems.

Another study [6] studied the impact of adding spatial cues, such as individual views and gaze awareness, to videoconferencing systems. They found that the spatial interfaces scored higher than a standard 2D control interface on social presence and co-presence measures but lower on task performance because of higher mental load.

The porta-person [20] is a telepresence device designed to improve the user experience of remote participants in hybrid meetings. It was inspired by the Hydra system [14] that uses a set-up of separate video displays representing each remote participant in order to preserve the notion of physical location of participants. The porta-person device contains a display screen, a video camera, stereo speakers and microphones on a rotating platform. It takes video images of the room and the device can present a video image of the remote participant and his or her voice. The device also conveys gaze direction and is designed to enhance the sense of social presence of remote meeting participants. The first experiences with the system were positive but the system turned out to be far too expensive to run field trials on a bigger scale.

3 Experimental Conditions and Hypotheses

The present study was designed to examine the effects of two different videoconferencing environments. We compare a standard 2D videoconferencing interface with an interface where video streams were presented to remote participants in an integrated 3D environment. The main difference between these environments is that the integrated 3D environment aims to support selective gaze and selective listening. In that environment gaze behavior of participants is transmitted in such a way that it is visible to all the participants (including the remote participant) to whom they look and who is visually attending to them.

3.1 Standard 2D Videoconferencing Interface

In the condition with the standard 2D videoconferencing interface (STANDARD), the three co-located participants (LP1, LP2 and LP3) are presented to the remote participant (RP) in a way similar to the presentation in a Skype or Adobe Connect meeting: the co-located participants have a webcam right in front of them. The camera images are presented to the remote participant in separate video frames positioned in a horizontal row and in 'random' order. Consequently, view directions on the screen do not match the real view directions. The images were presented to the RP on a classical large (52") video screen, see Figure 1.

Fig. 1. Presentation of co-located participants to the remote participant in STANDARD

3.2 Integrated 3D Videoconferencing Interface

In the condition with the integrated 3D video conferencing interface (3D) the placement of the cameras is different. See Figure 2 for an overview of the room with the camera settings. If the co-located participants look to the screen with the remote participant they look into the camera. In this condition view directions on the screen of the RP match the real view directions as good as possible.

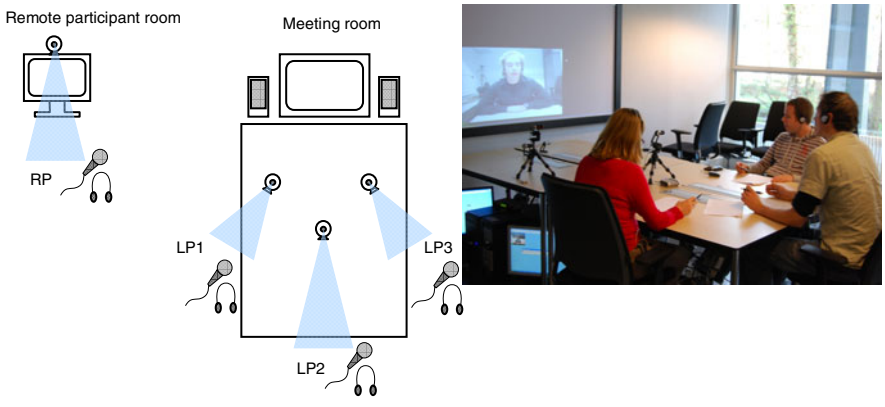

Fig. 2. Meeting room setting Integrated 3D version

Additionally the video images of the three co-located participants are presented to the remote participant in a more immersive representation around a virtual table, thus aiming to enhance the social presence of the remote participant. For this presentation we used the same video screen as was used in the other condition. Consequently the presentation was not really 3D but it offered the right perspective on the people in the meeting room. See Figure 3. Presentation of the remote participant in the meeting room was the same as in the other condition (see Figure 2, right picture).

3.3 Hypotheses

In both conditions we used the same cameras and the same screens for presentation of the video images. The quality of the images was good enough to capture non-verbal signals like facial expressions, eye movements and postures of participants in both conditions. However, recognition of gaze direction will be more difficult in STANDARD than in the 3D environment that was designed to reflect gaze directions in a natural way. Note that in STANDARD the co-located participants LP1 and LP3 look away from the camera when they look to the screen with the RP, hence in the perception of the RP they look away from him/her. Because of this, added to the fact that in STANDARD video images are presented to the RP in random order, we expect that in STANDARD (compared to 3D) the group process and turn-taking process will be impaired, resulting in lower scores on the group process questions and turn-taking, usability and recognition questions of the questionnaire described in section 4.4.

In addition, because in the 3D environment the co-located participants are presented to the remote participant in a more immersive way and they really seem to look at the remote participant when addressing him or her, the 3D condition is expected to establish a higher involvement of the remote participant in the group discussion (higher participation, a feeling of 'being there'). Hence, compared to STANDARD, the perceived social presence in the 3D condition is expected to be higher. The differences between STANDARD and 3D are expected to have more impact on the remote participants than on the participants in the meeting room.

Fig. 3. Remote participant room in the Integrated 3D version

4 User Study – Method

The two user environments for video-mediated hybrid meetings have been compared in a user study that measured the effects of the different environments on perceived social presence, satisfaction with the decision making process of the group and perceived turn-taking behavior and usability. In addition gaze behavior was observed. This section describes the setup of this study.

4.1 Participants and Experimental Design

Participants in the study were 40 young adults (5 women and 35 men) with ages ranging from 18 to 38 (most between 21 and 29). They were researchers (most PhD students) and students from the Computer Science department of the University of Twente who were not paid for their participation. They discussed in the hybrid meeting environment

for the first time. Participants took part in hybrid small group meetings (10 groups with four participants in each group). Three participants of each group met in a common (instrumented) meeting room and one participant took part remotely, via a videoconferencing system.

Within-group design is chosen for this experimental study, which means that each group performed a task in each of the two conditions. The conditions were counterbalanced, hence 5 groups started with the standard 2D videoconferencing interface and the other 5 started with the integrated 3D interface.

The experiment took place in the Smart XP Lab at the University of Twente. All sessions were captured with 4 web-cameras (1 per participant), 3 ceiling-mounted video cameras in the meeting room and one camera in the remote participant's room, capturing the image on the video screen the RP saw.

4.2 Group Decision Tasks

Since decision making tasks require more coordination and group member interaction than many other tasks [1], the groups were given two decision making tasks on which to come to consensus. One task was to select one student (out of three) to admit into the university's undergraduate program. The other task was to select a location (out of three possible locations) for a new 24-hour supermarket. These tasks were taken from [3] and adapted in the sense that there was no demonstrably best answer. According to Stasser and Steward [15] this is a judgement task and the best the group can do is come to consensus. In this kind of tasks the decision process is not so much focusing on exchanging critical information and finding the truth. Instead the decision process "is more aptly characterized by egalitarian social combination schemes such as majority- or plurality-wins models" [15; pp. 432].

In a few additional adaptations to the tasks we followed [8]. Instead of receiving different hidden profiles, all group members had the same information about the student candidates and the possible locations. This was done to avoid participants looking at the paper description during the discussion. As our intention was to observe the visual attention we took away the paper descriptions during discussion. To initiate an engaging group discussion, the participants received different roles in the discussion: they had to defend different beliefs and values probably leading to different choices. E.g., for the student selection task one participant role emphasized intellectual ability while another role emphasized diversity in cultural backgrounds.

The tasks were counter-balanced within each condition and order of condition, to rule out influences of the tasks on the results.

4.3 Procedure

Participants were scheduled in groups of 4 on the basis of availability at certain times. In some groups participants knew each other, in other groups they did not and there were mixed groups as well. Participants met in a room next to the meeting room. They received a short introduction, only stating that they participated in a user study on videoconferencing support for group meetings and that they would engage in two group discussions, each followed by filling in a questionnaire. Then the remote participant was randomly chosen from the four participants in the group and brought to a separate room, while the three other group members entered the meeting room.

The group started with a warm-up discussion of 5 minutes about a topic they chose from a list of topics. During this discussion we checked if all the equipment functioned well. Then participants got 5 minutes time to independently read the first group decision task. They studied the available alternatives and their role in the discussion, and they were asked to make a preliminary choice. They were told in advance that the task descriptions would not be available during the discussion. Additional time was given on request. After the task descriptions were taken away, participants engaged in a discussion for 15 minutes. Two minutes before the end of the discussion time the experimenter warned the participants they only had two minutes left to come to a final decision which every team member can agree with. After the discussion all four participants went to the room where they met, to fill in a questionnaire. In the meantime settings in the meeting room and the remote participant room where prepared for the second part of the session.

When people returned in the meeting room (respectively the remote participant room) they received the second group decision task and followed the same procedure in the other condition: 5 minutes reading, handing in task descriptions, 15 minutes discussion - with a warning to come to consensus - and filling in a questionnaire in the other room. In the end there was a short post-interview with the group.

4.4 Measures

The questionnaire participants filled in after each group decision task consisted of several parts: a part to assess perceived group process (18 questions), a part with a social presence questionnaire (19 questions), and a part with 11 questions about usability, turn-taking and recognition of gaze direction and other non-verbal signals. All questions were rated on a 5-point Likert scale, where '1' meant 'Strongly disagree' and '5' meant 'Strongly agree'.

The group process part of the questionnaire consisted of 12 questions about perceived group process quality [10], 5 questions about satisfaction with the decision making process [11, 12] and one question about overall satisfaction with the final group decision [12].

In the social presence questionnaire we included parts of the validated social presence questionnaire of Harms and Biocca [5] and a few questions taken from Hauber et al. [6]. From Harms and Biocca we used 16 questions: the complete subscales Co-presence and Attention Allocation and a few items from the subscales Message Understanding and Perceived Behavioral Interdependence. We left out the subscales Perceived Emotional Interdependence and Perceived Affective Understanding because we expected these to be of less relevance in this meeting context. The three questions taken from Hauber et al. [6] were labelled Co-presence as well. We added them because they were formulated in relation to face-to-face contact (e.g., "sometimes it was just like being face-to-face with the RP/LPs" and "it sometimes felt as if the RP/LPs and I were in the same room") and hence were expected to be very relevant for our study. The co-presence questions of Harms and Biocca were formulated in terms of noticing each other (e.g., "The RP/LPs always noticed me" and "My presence was obvious to the RP/LPs"), which of course is also relevant.

The rest of the questions were about turn-taking, floor control and usability, taken from [6, 12] and a few questions we made for this study: about recognition of gaze direction and other non-verbal signals.

During the discussions visual focus of attention of the participants was observed and annotated real-time. Every observer monitored one of the participants and annotated who the participant was looking at. If the participant did not look at one of the other participants it was annotated the participant looked somewhere else.

5 Results

This section presents the results of the questionnaires, an analysis of the observed gaze behavior of the participants, and the results of the post-interviews.

5.1 Group Process and Satisfaction with the Outcome

We used exploratory factor analysis (principal component analysis with oblimin rotation) to find underlying dimensions in the data from the questionnaire part on perceived group process quality and satisfaction with the decision making process. Two questions about overall satisfaction with the process and the outcome were excluded from this analysis, as well as a question about trustworthiness of the group members. Because of their deviating form, these questions were studied separately. In the factor analysis the question "group members brought a variety of perspectives to bear on the tasks" loaded on a separate factor and will be treated separately as well.

The remaining questions loaded on three factors. The first factor contains 7 questions (e.g., "The general quality of the group members' contributions to group discussions was very good" and "The evaluation of arguments was very thorough") and was labelled Group process and contributions. The second factor contains 4 questions (e.g., "The group discussions were unorganized" and "There were disruptive conflicts") and was labelled Organization and cooperation. The third factor contains 3 questions (e.g., "People were friendly in my group" and "Comments reflected respect for one another") and was labelled Group members.

Cronbach Alpha tests were used to analyze the reliability of the subscales identified by the factor analysis for both the STANDARD and the 3D condition. For *Group process and contributions* the alpha reliabilities were .73 in 3D and .85 in STANDARD. *Organization and cooperation* had an alpha of .66 in 3D and .75 in STANDARD and the alphas of *Group member* were .65 (3D) and .89 (STANDARD). Hence reliabilities varied from reasonable (.65) to high (.89) [4].

We used Wilcoxon signed ranks tests (α=.05) to analyze the differences between the two experimental conditions on the three subscales and four questions of the group process part of the questionnaire. We did the tests for all participants (40 persons) and for the 10 remote participants (RP) and 30 co-located participants (LPs) separately. The results of the analyses are shown in Table 1.

As can be seen in Table 1, there were no factors or questions on which the STANDARD condition scored significantly higher than the 3D condition. 3D scored significantly better on Organization and cooperation. This effect is strong for the LPs ($p < .01$) and not significant for the remote participants. Here we have to keep in mind that

the group of remote participants was only small (10 people) and hence finding statistically significant differences will only be possible if the differences are really consistent and quite large. We found a marginally significant difference in favor of 3D for Group member (all group members) and Group process and contribution (only RPs). Furthermore, to our surprise, with the RPs we found a significant difference, in favor of 3D, on satisfaction with the final decision. On the questions on trustworthiness of group members and satisfaction with the solution process there were no significant difference between the conditions.

Table 1. Differences between STANDARD and 3D on Group Process and Satisfaction with the Outcome. Columns "Best" shows the condition (3D or ST) that scored significantly higher.

Factor or question	All		RP		LPs	
	Best	Z	Best	Z	Best	Z
Group process and contribution			3D	-1.79 †		
Organization and cooperation	3D	-2.37 *			3D	-2.88 **
Group members	3D	-1.80 †				
Variety of perspectives					3D	-1.97 *
Trustworthiness group members						
Satisfaction solution process						
Satisfaction final decision			3D	-1.98 *		

† $p < 0.1$ * $p < 0.05$ ** $p < 0.01$

5.2 Social Presence

Cronbach Alpha tests were used to find out if the subscales of Harms and Biocca and Hauber were reliable in both the STANDARD and the 3D condition. The results are shown in Tabel 2.

Table 2. Cronbach's Alphas for the social presence subscales of Harms and Biocca and the co-presence subscale of Hauber et al

Factor	3D	STANDARD
Co-Presence Harms and Biocca (6 items)	.76	.82
Attention Allocation (6 items)	.57	.51
Message Understanding (2 items)	.72	.70
Perceived Behavioral Interdependence (2 items)	.66	.73
Co-Presence Hauber et al. (3 items)	.79	.80

Except for Attention Allocation, all scales are reliable hence we decided to use the scales in the analyses. To study the differential effects of the two experimental conditions on social presence we again used the Wilcoxon signed rank test. The results are

presented in Table 3. Because analyses for all participants (All) and for the remote participant (RP) did not show significant differences, these columns are left empty. During the interviews we noticed there was a difference in perceptions between the co-located participants LP2 that were in the position facing the screen with the remote participant and LP1 and LP3 that had to look to their left or right to see the remote participant. Hence we repeated the Wilcoxon tests for LP2 (10 persons) and LP1+LP3 (20 persons) separately as well.

Table 3. Differences between STANDARD and 3D on Social Presence. Columns "Best" shows the condition (3D or ST) that scored significantly higher.

Factor	All	RP	LPs		LP1+LP3		LP2
			Best	Z	Best	Z	
Co-Presence Harms and Biocca					3D	-2.04*	
Attention Allocation			3D	-1.66†	3D	-1.99*	
Message Understanding							
Perceived Behavioral Interdep.			3D	-1.80†	3D	-3.14**	
Co-Presence Hauber et al.							

† $p < 0.1$ * $p < 0.05$ ** $p < 0.01$

For "All participants" and for the "Remote Participant" and the participants on location 2 no significant differences between the conditions were found on any of the social presence subscales. Participants on locations 1 and 3, however, did perceive significant differences, all in favor of 3D, on three of the subscales of the social presence questionnaire of Harms and Biocca [5]: Co-presence, Attention Allocation and Perceived Behavioral Interdependance.

5.3 Turn-Taking, Usability, Recognition of Non-verbal Cues and Gaze Direction

The remaining questions on turn-taking, usability and recognition of non-verbal signals and gaze direction were analyzed separately because no reliable subscales could be identified. The questions and the results of Wilcoxon tests we used to analyze the differences between the experimental conditions can be found in Table 4.

No significant differences between the conditions were found for any of the turn-taking questions. The perceived effort it took to follow the discussion and ease of recognition of non-verbal signals did not differ significantly either. But very significant differences, again in favor of 3D, were found for all participants on the statements "I got the feeling that the other participants/the remote participant looked at me " and "It was clear to whom the other participants/the remote participants talked." Hence the participants clearly noticed the intended difference between the two conditions.

Table 4. Differences between STANDARD and 3D on Turn-taking, Usability, Recognition of Gaze direction and Non-verbal Cues. Columns "Best" shows the condition (3D or ST) that scored significantly higher.

Question	All		RP		LPs	
	Best	Z	Best	Z	Best	Z
I knew exactly when it was my turn to speak						
We were never talking over one another						
There was a lot of time when no-one spoke at all						
I could always clearly hear the voices of the other group members (LPs)			3D	-1.89 †		
It was easy to take my speaker turn when I wished to do so						
It took me a lot of effort to follow the discussion						
I could recognize non-verbal signals of the RP/LPs easily						
I got the feeling that the RP/LPs looked at me	3D	-3.00**	3D	-2.10*	3D	-2.18*
It was clear to whom the RP/LPs talked	3D	-3.49***	3D	-2.26*	3D	-2.63**
The presentation of the RP/LPs on the screen was appealing	3D	-1.90 †	3D	-1.90 †		

† $p < 0.1$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

5.4 Visual Focus of Attention

To find out if there was a difference between the experimental conditions in the frequency and duration the co-located participants looked at the remote participant we used the annotations of the observers. For every participant in the discussions, we derived the number of times they looked at each of the other participants in the discussions. From the durations of each of these counted gaze acts we also derived how long (in seconds) the participant looked at each of the other participants during the whole discussion. Because there were small differences between the durations of the discussions, in the analyses we used variables that correct for duration of the discussion. Instead of using the number of times participant x looked at participant y during the discussion we used the number of times participant x would have looked at participant y if the discussion would have lasted an hour (relative number of times). Instead of using the total time participant x looked at participant y during the discussion we used the percentage of the discussion time participant x looked at participant y (percentage of time).

To find out if the two experimental conditions were different in how often and how long the co-located participants looked at the remote participant we used the paired t-test. The results are presented in Table 5.

No significant differences were found between the conditions in how often and how long the co-located participants looked at the remote participant. Similar analyses of the relative number of times and the percentage of time other participants in the discussion (LP1, LP2, LP3) were looked at yielded no significant differences between STANDARD and 3D either.

Table 5. Differences between STANDARD and 3D in Gaze Behavior

	STANDARD		3D		Results paired t-test		
	Mean	SD	Mean	SD	t	df	*p*
Relative number of times LPs looked at RP	191.4	95.7	188.7	72.4	.18	29	.86
Percentage of time LPs looked at RP	15.9	8.13	18.9	9.4	-1.54	29	.13

5.5 Group Interviews

The post-interviews with the groups were open-ended, starting with the question if they had a preference for one of the environments and, if so, what was the preferred environment and what were the reasons. Seven groups unanimously chose 3D as their favorite. In the three remaining groups only the remote participants had a deviating opinion: one of them said he had no preference, the other two preferred STANDARD. They did not like the 3D view with the virtual table and the backgrounds of the images that did not fit together very well. Most important reasons that were mentioned: 3D was more natural (mentioned by 7 groups), more intuitive (mentioned twice). Some groups added that 3D worked well: they really felt this was a good way to meet or to have a conversation. In 3D it was more clear who looked/talked to whom (5 times) and most remote participants said that in 3D it was clear when people looked at them and they could see if they had the attention of the others. A few remote participants said they had the idea they looked the others in the eyes. In STANDARD they had to find out first if they were addressed and if it was their turn to speak. Moreover, in STANDARD many remote participants did not know to whom the co-located participants talked. Four groups mentioned that in 3D the remote participant was more involved in the group and in two groups the co-located participants said that in STANDARD they paid less attention to the remote participant. Participants on locations 1 and 3 (LP1 and LP3) often mentioned the cameras in STANDARD were inconvenient – too close in front of them. Another thing often mentioned was that in STANDARD it was difficult to look at the remote participant on the screen and at the same time look in the camera. Only people who were aware of how the remote participant would probably see them mentioned this. Others reacted by saying they never thought about the fact that the remote participant would see them from the side if they looked at the image of the remote participant. Often participants LP2, the people right in front of the screen with the remote participant, said they did not notice much difference between 3D and STANDARD.

6 Discussion

The participants clearly noticed the intended difference between the two conditions. In 3D both the remote participants and the co-located participants stated they could better distinguish who was being addressed. In addition, many of the remote participants got the feeling the co-located participants looked at them and many of the co-located participants got the feeling the remote participant looked at them. As a result of this we would expect the remote participants to be more involved in the 3D meetings than in the STANDARD meetings. However, no significant differences were found for the *remote* participants on any of the social presence subscales, though mean scores were consistently higher in 3D. (Hence the differential effects, if any, were not strong enough to be significant for 10 participants.) The situation was different for the two local co-located participants LP1 and LP3 that were affected by the differences between the two interfaces: here we did measure significant differences in favor of 3D on the social presence subscales co-presence, attention allocation and perceived behavioral interdependence. This indicates, for instance, that in the 3D condition LP1 and LP3 felt more present to and felt more noticed by the remote group member and visa versa and they remained more focused on each other during interaction than in the STANDARD condition.

To our surprise, no significant differences between the conditions were found for any of the turn-taking questions and the perceived effort it took to follow the discussion did not differ either. Hence, though the participants could distinguish gaze directions better, that did not influence their perceived turn-taking behavior. The analysis we did on the observer annotations of visual focus of attention did not show any significant differences between STANDARD and 3D in how often or how long the co-located participants looked at the remote participant. If the 3D condition would resemble face-to-face communication more than the STANDARD condition we would have expected more attention for the remote participant in the 3D condition. It will be interesting to do more analyses on the behavioral data to see if objective measurements support the outcomes of the subjective measurements and for instance number of interruptions, the time no-one spoke at all, or participants started talking at the same time really did not differ between the conditions.

On the questionnaires measuring the perceived quality of the group process and the discussions, significant differences in favor of 3D were found on the organization of and cooperation in the meetings and on the opinions about the other group members. Trustworthiness of the group members did not differ between the conditions.

The results of the group interviews at the end of the user study mainly support the results of the questionnaires: overall 3D was preferred by most participants. Actually we did expect the differences between the conditions to be very clear for remote participants and not so clear for the co-located participants because for them the only difference between the conditions was the gaze behavior of the RP. But obviously the differences were clearly noticeable, especially for LP1 and LP3. The 3D environment was found to be more natural and intuitive and suitable to support remote meetings or conversations. Based on the interviews we would have expected to find differences in perceived turn-talking behavior. Possible explanation for the absence of these differences might be that the meeting participants thought that in STANDARD they succeeded to have equally fluent turn-taking behavior, despite the missing directional

gaze cues. Further analyses of observational data should show if that was really the case or meeting participants are not really very conscious about their turn-taking behavior.

7 Conclusions

This study compared two user environments that aim to support remote participation in hybrid meeting. One environment, named STANDARD, is the control environment. It resembles a conventional video conferencing environment: co-located meeting participants all look straight into their webcam and their video images are presented to the remote participant in separate frames, presented in a horizontal row. This environment fails to support selective gaze and selective listening. The experimental environment, named integrated 3D or shortly 3D, is designed to convey gaze directions in a natural way. Results of the user study indicate that the 3D environment is promising: all differences that were significant were in favor of 3D and according to most participants the environment clearly supports selective gaze and selective listening in a natural way. Moreover, remote participants often mentioned they felt as if co-located participants looked at them. Nevertheless, on the social presence measures no significant differences between the two environments were found for remote participants, possibly because the number of groups and hence of remote participants was only 10. Another possible explanation was put forward by Hauber et al. [6]. In their study they did not find any significant results on social presence measures. They conclude that the social presence measure might not be sensible enough to find differences and suggest to add objective or physiological measurements. In our study we did find significantly higher scores for 3D on the social presence subscales co-presence, attention allocation, and perceived behavioral interdependence for co-located participants in locations 1 and 3, the locations where they had to look to their left or right to see the screen with the remote participant. In addition we found some significant differences on measures of perceived quality of group process. We did not find any significant differences between 3D and STANDARD in perceived turn-taking behavior. Considering the fact that in the interviews and the questionnaires the participants clearly stated that in 3D they could better distinguish who was being addressed, we intend to further analyze observational data (gaze annotations, video recordings) to find out if differences in turn-taking behavior occurred between the two environments.

Acknowledgments. The work reported in this paper is sponsored by the European IST Programme Project FP6-0033812 (AMIDA) and the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 231287 (SSPNet). 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. op den Akker, H.J.A., Hofs, D.H.W., Hondorp, G.H.W., op den Akker, H., Zwiers, J., Nijholt, A.: Supporting Engagement and Floor Control in Hybrid Meetings. In: Esposito, A., Vích, R. (eds.) Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions. LNCS, vol. 5641, pp. 276–290. Springer, Heidelberg (2009)

2. Burke, K., Aytes, K., Chidambaram, L., Johnson, J.: A Study of Partially Distributed Work Groups: The Impact of Media, Location, and Time on Perceptions and Performance. Small Group Research 30(4), 453–490 (1999)
3. DiMicco, J., Pandolfo, A., Bender, W.: Influencing Group Participation with a Shared Display. In: Proc. CSCW 2004, pp. 614–623. ACM Press, New York (2004)
4. Field, A.: Discovering Statistics Using SPSS, 3rd edn. Sage Publications Ltd, Thousand Oaks (2009)
5. Harms, C., Biocca, F.: Internal Consistency and Reliability of the Networked Minds Social Presence Measure. In: Alcaniz, M., Rey, B. (eds.) Seventh Annual International Workshop: Presence 2004, pp. 246–251. Universidad Politecnica de Valencia, Valencia (2004)
6. Hauber, J., Regenbrecht, H., Billinghurst, M., Cockburn, A.: Spatiality in videoconferencing: trade-offs between efficiency and social presence. In: Proc. CSCW 2006, pp. 413–422. ACM Press, New York (2006)
7. Kendon, A.: Some Functions of Gaze-Direction in Social Interaction. Acta Psychologica 26, 22–63 (1967)
8. Kulyk, O., Wang, C., Terken, J.: Real-Time Feedback Based on Nonverbal Behaviour to Enhance Social Dynamics in Small Group Meetings. In: Renals, S., Bengio, S. (eds.) MLMI 2005. LNCS, vol. 3869, pp. 150–161. Springer, Heidelberg (2006)
9. Nijholt, A., Rienks, R.J., Zwiers, J., Reidsma, D.: Online and Off-line Visualization of Meeting Information and Meeting Support. The Visual Computer 22(12), 965–976 (2006)
10. Olaniran, B.A.: A Model of Group Satisfaction in Computer Mediated Communication and Face-to-Face Meetings. Behaviour & Information Technology 15(1), 24–36 (1996)
11. Paul, S., Seetharaman, P., Ramamurthy, K.: User Satisfaction with System, Decision Process, and Outcome in GDSS Based Meeting: An Experimental Investigation. In: Proc. of the 37th Hawaii International Conference on System Sciences. IEEE Computer Society Press, Los Alamitos (2004)
12. Post, W., Elling, E., Cremers, A., Kraaij, W.: Experimental Comparison of Multimodal Meeting Browsers. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4558, pp. 118–127. Springer, Heidelberg (2007)
13. Sellen, A.J.: Remote Conversations: The Effects of Mediating Talk With Technology. Human–Computer Interaction 10, 401–444 (1995)
14. Sellen, A., Buxton, B., Arnott, J.: Using spatial cues to improve videoconferencing. In: Proc. CHI 1992, pp. 651–652. ACM Press, New York (1992)
15. Stasser, G., Stewart, D.: Discovery of Hidden Profiles by Decision-Making Groups: Solving a Problem Versus Making a Judgment. Journal of Personality and Social Psychology 63(3), 426–434 (1992)
16. Vertegaal, R.: The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. In: Proc. CHI 1999, pp. 294–301. ACM Press, New York (1999)
17. Vinciarelli, A., Pantic, M., Bourlard, H.: Social Signal Processing: Survey of an Emerging Domain. Image and Vision Computing 27(12), 1743–1759 (2009)
18. Whittaker, S.: Rethinking Video as a Technology for Interpersonal Communications: Theory and Design Implications. Intl. J. of Man-Machine Studies 42, 50–529 (1995)
19. Whittaker, S.: Theories and Methods in Mediated Communication. In: Handbook of Discourse Processes, Erlbaum, NJ, pp. 243–286 (2002)
20. Yankelovich, N., Simpson, N., Kaplan, J., Provino, J.: Porta-Person: Telepresence for the Connected Conference Room. In: Proc. CHI 2007, pp. 2789–2794. ACM Press, New York (2007)