GColl Group-to-group Videoconferencing System: Design and First Experiences

Petr Slovák
CESNET z. s. p. o.
Zikova 4, 160 00 Prague
Masaryk University
Botanická 68a, 602 00 Brno
Czech Republic
Email: slovak@ics.muni.cz

Pavel Troubil
CESNET z. s. p. o.
Zikova 4, 160 00 Prague
Masaryk University
Botanická 68a, 602 00 Brno
Czech Republic
Email: pavel@ics.muni.cz

Petr Holub
CESNET z. s. p. o.
Zikova 4, 160 00 Prague
Masaryk University
Botanická 68a, 602 00 Brno
Czech Republic
Email: hopet@ics.muni.cz

Abstract—**We present a videoconferencing tool, GColl, which aims to support collaboration among remote *groups* of participants. GColl supports mutual gaze as well as partial gaze awareness for all participants, while still retaining a modest technical requirements: a camera and an echo-canceling microphone at each site; and a laptop with two USB cameras for each user. The environment is also easily deployed and allows quick changes in numbers of participants at individual sites. It is therefore suitable even for ad-hoc groups or teams with small budgets. A quantitative user study has been conducted in order to evaluate functionality of GColl with promising results. Additionally, the tool is available for download as an open-source project.**

I. I

Videoconferencing systems and tools are used quite often in commercial organizations as well as academic institutions to support teams with remote participants. Currently, there is a great number of tools available for videoconferencing among individuals (such as Skype, Adobe Connect etc.), but much less work has been done in creating environments that would support collaboration among remote *groups of people*. At the same time, a simple extension of the existing tools is not ideal as the problems known from previous research (e.g., preservation of gaze awareness and other non-verbal cues) become even more salient in the group-to-group setting.

Teams that use videoconferencing tools are of various types: from those, that have a very stable task and team structure, to teams that are created only on a short term basis or require adaptability to frequent changes in the number of participants and/or changes in videoconferencing locations. While the first might settle for a fixed videoconferencing system in a dedicated room, such environments are not suitable for the latter type, which we will denote in this paper as *ad-hoc groups*.

Although there are some academic videoconferencing systems, which have been designed directly for multi-person communication, e.g., [21], [14], [17], [13], none is completely suitable for ad-hoc groups. For example, if a size of a group increases at one site, a non-trivial change to the physical layout at remote sides might be needed. Also, it might not be easy to transfer the videoconferencing system into a different room or another building as it usually comprises of a complex structure of cameras and viewing screens. In other words, once set up, these designs provide the users with excellent communication environment; however, ad-hoc groups might be forced to spend a lot of energy on re-installation or transport of the system.

On the other hand, most of the widely available commercial solutions (such as H.323/SIP systems) are based on a simple extension of the individual videoconferencing concept – each communicating group is equipped with a large LCD with a camera and some kind of a echo-canceling microphone. Most of the non-verbal signals such as gaze awareness information are therefore lost in these systems.

In this paper, we present our videoconferencing environment, GColl, which is based on a compromise between the need for preserving the non-verbal cues present in face-to-face communication and the requirements of low-cost and flexibility. While keeping all the aspects important for ad-hoc groups (easy mobility, minimal installation, the possibility to easily change the number of participants even during a conference), we are also able to transfer a reasonable portion of non-verbal signals such as mutual gaze and partial gaze awareness (the information of who is looking at you). The technical requirements of GColl are also modest: it requires only a camera and an echo-canceling microphone at each site, and a personal laptop with two web cameras for each user.

We have also conducted a medium-scale quantitative user-study (90 participants) to evaluate functionality of GColl, by comparing it to the face-to-face environment and an environment similar to common commercial systems.

In the following sections, we first give an overview of related work and describe briefly the basic concept of our design, which was fully specified in [18]. We then present the details of our current implementation as well as discuss conducted user study.

II. R W

Concept of *gaze awareness* has been studied in great detail in the literature due to its importance for effective communication as well as other task-related activities (e.g., in [20], [12]). A widely accepted definition of gaze awareness, proposed by Monk and Gale [12], distinguishes among three forms: *full*

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8344
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8344

gaze awareness – knowledge of the current object in someone else's visual attention; *partial gaze awareness* – being aware of the general direction someone is looking (e.g., whether he/she is looking at you, or at someone else); and *mutual gaze* – possibility of eye contact.

In most videoconferencing environments, gaze information is not conveyed easily due to the usual discrepancy between the camera position and the place of visualization of the other person's eyes. While several videoconferencing systems were invented to mediate some or all forms of gaze awareness (e.g., GAZE2 [21], MAJIC [14], Hydra [17] and Multiview [13]), none of these are, however, directly suitable for ad-hoc groups due to either the lack of support for group-to-group interaction ([21], [14], [17]), or problems with mobility and flexibility [13].

Another line of research has recently focused on the effects of *mixed presence* (i.e., a collaboration among multiple distributed sites, each with a co-located group, which are connected by a communication channel; thus allowing the use of both face-to-face and computer mediated communication) where the problem of *presence disparity* was identified by several studies (e.g., [4], [19], [2]). This term describes the tendency of the users at each individual site to form a strong sense of an in-group, which in turn leads to the users collaborating mainly with other co-located users and neglecting those, who are physically remote. Therefore, presence disparity is an undesirable effect which an ideal environment should mitigate.

III. D GC E

The environment was designed with focus on allowing mutual gaze and partial gaze awareness, while retaining flexibility and mobility required by ad-hoc groups. More detailed description of the environment is subject of [18].

A. Environment Setup

a) Components: The environment comprises a camera and an audio system for the whole group, as well as a personal computer — typically laptop computer — with two webcams attached for each user, as shown in Figure 1. The personal computers are needed to have individual video capture and playback capabilities, for otherwise all the users would "share the same eyes" through the group camera and a projection screen. We believe the assumption of personal computer is not too restrictive for anticipated users, be it in commercial sphere or academia. Furthermore, the proposed two camera setup can be built into the personal computers in a similar way it is frequently implemented now with a single camera.

b) Audio: All participants at each site share the audio by default. Audio is captured by a group microphone or downmixed microphone array and sound is played by speakers for the whole group. Therefore, some kind of echo cancellation is also necessary — be it in a microphone with echo cancellation (e.g., ClearOne AccuMic) or dedicated echo-canceling device (e.g., Polycom SoundStructure).

The group audio can be substituted by personal monaural headsets with short-range microphones, to be able to work with individual audio streams. Having the headset attached to one ear only allows for full on-site participation of individuals in the group. For more natural operation, wireless headsets are required, but applicability of these is often severely limited by their low sound fidelity.

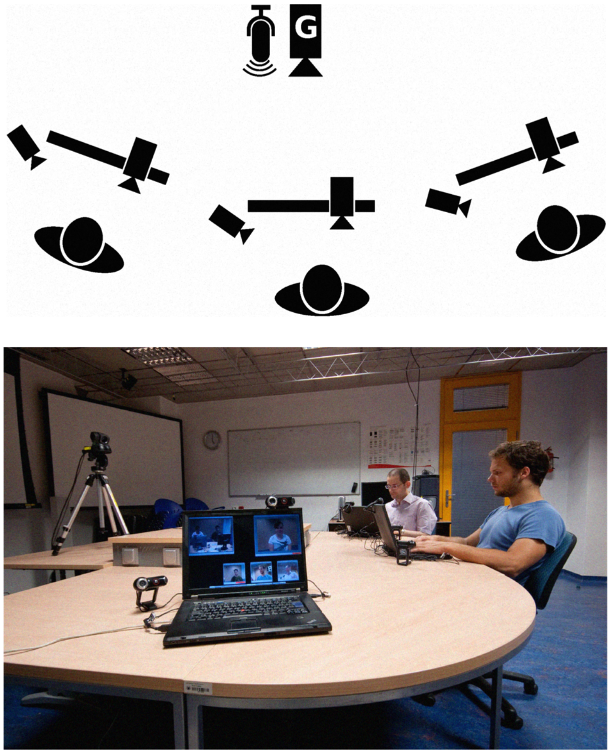

Fig. 1. A scheme of site installation and a photo of a site running the videoconference.

c) Video: As mentioned above, there are three video sources for each user: (1) the *group camera*, which provides video of the whole group, (2) a *focus camera* attached to the top-right of the screen, and (3) a *side camera*, positioned typically few inches from the bottom-left edge of the screen.

The screen with video playback is divided into three parts: (1) *whole group* video stream at the top left of the screen (with group videos possibly merged in case of more than two groups participating), (2) *focus* window at the top right of the screen, and (3) individual video streams at the bottom of the screen. When a user focuses on another user, he looks at the top-right section of the screen, where the focus camera is also positioned. The overall setup of the desktop with focus camera attached is shown in Figure 2, as well as a photo from one of our test groups.

This setup enables mediation of mutual gaze sensation. It utilizes findings by Chen [5] that show asymmetry in human sensitivity to eye contact: people would still perceive eye contact if the other person's gaze is directed less than 5° below the camera. By having a fixed place for the focus window in GColl, it is easy to attach the focus camera appropriately

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8344
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8344

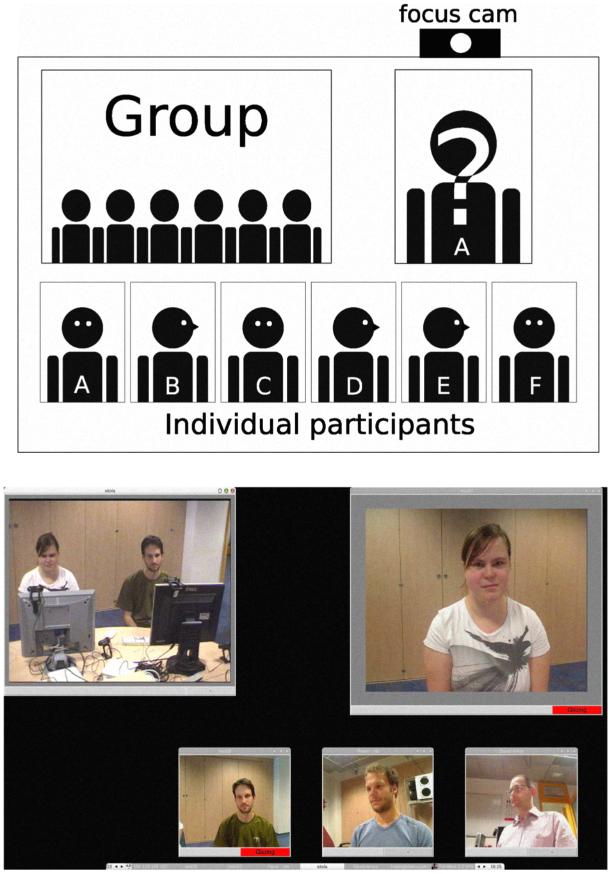

Fig. 2. First image depicts a scheme of the visualization screen layout. The respective user is being focused by participants A, C, and F, while himself focusing on user A. The other participants are focused at someone else (not knowing who exactly). On the photo, the user is focused at another participant, who is focused at him, too. There's one another participant focusing at the respective user – both of the focuses can be distinguished both by perceived eye-contact and visual highlight (red bar).

close to the visualized eyes[1]. The focus part of the GColl is located at the top right of the screen as described above, but the approach is general enough so that the focus window can be re-positioned along the top edge of the screen, while preserving the 5° vertical range, provided the side camera is far enough that the image from it is naturally distinguished.

B. Operations

When the GColl environments starts, each user is provided with the group video stream and individual video streams at the bottom of the screen. After clicking at any of individual video stream windows, the respective user becomes in focus, i.e., (1) the selected video stream is shown in enlarged window in the focus area below the focus camera, and (2) the user, that is being focused on, starts to receive the video from the focus camera of the user who initiated the focus action and the respective window is slightly visually highlighted. Since only one user can be in focus at a time, any previously selected user is automatically deselected, i.e., shown in small window and receiving side camera stream again. Clicking on the group camera window removes focus from the previously selected user without selecting anyone else. Thus, by looking at the individual video streams at the bottom of the screen, each user can distinguish, in whose focus he is, depending on whether side camera stream without eye contact or focus camera stream with eye contact is received. This process is very natural due to the strong human perception of eye contact. Taking the example shown in scheme in Figure 2, the user focuses on user A, while he knows that he is also in focus of users A, C, and F. Of course, it is a slightly more restricted version of mutual gaze than the one provided by Multiview, but it is a price we pay for the flexibility and mobility of GColl environment as a whole.

In default operation, the group audio is only used, providing no focus cues. With individual audio streams, the personalized audio playback can be used for additional focus cues. When the participant becomes in focus of some other participant, the audio level between those two participants is is increased by approximately 15 %. A similar technique is used by GAZE-2 system to ease management of side conversations.

IV. GC I

A. Data Distribution

The design of GColl requires at least one videostream being transmitted between each couple of videoconference participants in both directions. Consequently, multipoint-to-multipoint transmission must be established to set up the GColl environment for multiple remote participants.

Multicast might be used to support multipoint videoconferences at the network level. Unfortunately, it is not widely available and, in our case, each client workstation would have to receive both camera streams from all other participants; videoconferencing clients would need a further logic in order to process signaling from other clients and choose the right videostreams to display.

Due to these problems, we have decided to base GColl on RUM2 [9] — a modular packet reflector that was developed in our laboratory. Packet reflector is a server application, which implements multicast functionality at the application level of network stack by sending each incoming packet to all "connected" clients. Lower scalability of this solution in comparison to multicast is not relevant for our videoconferencing environment, since GColl aims to support groups up to 10–15 members.

RUM2 is also very advantageous in terms of data processing. While multicast delivers exactly the same data as clients send, packet reflector allows for arbitrary data processing by means of pluggable modules at the server side.

These capabilities were exploited to implement gaze awareness functionality in GColl, by using a single module. Each client sends videostreams from both cameras to the packet reflector. The GColl module then chooses one of the streams to be sent to each of other videoconference participants: *focus* camera stream is sent only to the user in the sender's focus

[1] For example, if eye of the camera is placed 5-6 cm from the visualized eyes (which was the distance maintained in our evaluation study), it suffices if the user sits approximately 60-70 cm from the screen.

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8344
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8344

window (if any); *side* camera stream is sent to all others. Similarly, the users don't receive streams of their own group camera, as it would not be displayed anyway. By mangling RTP headers, personal videostreams are managed in a manner that makes their switching transparent for client tools; client also sends full frames of both streams upon focus switch to avoid interleaving of the images at remote participants' screens. When personal audio is used, the module sends each client audiostreams only from the remote participants, avoiding "echoes" caused by the latency of network and audio system: the user would hear all words twice – locally and then delayed from his/her headset.

Bandwidth Saving: By sending the needed video data to individual recipients only, our approach to data distribution saves a significant amount of bandwidth at participants' workstations: when one multicast group for all personal webcams is used, required downstream bandwidth might be expressed as

$$\sum_{p\in P}(bf_p + bs_p) + \sum_{s\in S_r} b_s,$$

where P denotes the set of all participants excluding the receiving one, S_r is the set of all sites *remote* to the receiver, and bf_p, bs_p, b_s denote bitrates of received streams — user's front webcam, user's side webcam and group camera respectively. Let's denote "one of bf_p, bs_p" as b_p; GColl with packet reflector based data distribution then requires only

$$\sum_{p\in P} b_p + \sum_{s\in S_r} b_s$$

of downstream bandwidth. Under simplifying assumption of all videostreams being equal in bitrate, the proportion of bandwidth saved by GColl equals to

$$1 - \frac{|P| + |S_r|}{2 * |P| + |S_r|},$$

which makes 46 % for six users at two sites. Also, if more than one packet reflector is available (e.g., one at each site), further data distribution optimizations are possible — especially on the links between the remote sites rather than at participants' workstations.

If there was a separate multicast group for each of users' webcams (i.e., two groups per user), multicast would require the same bandwidth as packet reflector does. Upon switching focus by any participant, other two would then leave the front-camera or side-camera multicast group and join the other one. However, such a solution is not an option for GColl due to high latency of joining and leaving multicast groups.

If the personal audio was used, the packet reflector also saves slight amount of the downstream bandwidth by blocking some of the audio streams. Due to low bitrate of audio streams in comparison to the video ones, this reduction is far less significant than blocking of the video streams. Unfortunately, upstream bandwidth cannot be saved at participants' workstations, since each user is focused at someone else most of the time, which calls for permanent transmission of both camera streams.

The packet reflector functionality may also be used for a definition of client machines, that are permitted to join GColl videoconference, authentication of participants, distribution of some additional data, etc.

B. Client Implementation

We have used Vic videoconferencing tool (a part of UCL Media Tools[2], version 1.4.0beta – VicH264, latest snapshot of "common" libraries for 64bit compatibility) as a base for the client-side implementation of GColl design. A number of new functions was implemented so that Vic could be used as the video transmission tool in the GColl environment. These comprise especially automated layout of the videostream windows on user display, switching streams in the *focus* window, and client-reflector communication. Since original Vic can capture and transmit video from a single source only, we had to implement support for two cameras required by the GColl design.

Although the GColl design has no limitation on number of sites participating in the conference, the client now supports displaying of single remote whole group stream only. We are currently implementing the support of more sites – all remote *group camera* streams will be merged (as a single tiled image) into the *whole group* frame. All necessary logic for the composition will be done by the packet reflector, thus eliminating any computational load put on the client workstations.

Originally, personal audio was handled separately by a RAT audio client (also a part of the UCL Media Tools), but in the current version, RAT has been incorporated into the GColl client to simplify user operation and communication between audio and video tools, and to allow future extensions.

The GColl client as well as packet reflector are freely available for download at [1].

C. Client-Reflector Communication

GColl clients communicate with the packet reflector over RAP — Reflector Administration Protocol [7]. The signalization comprises particularly the mode of the client (personal or group one), identification of the site where the client is located, giving semantics to the video streams (i.e., telling the reflector which of the two streams transmitted by the client comes from focus/side camera) and notifications when the user changes his focus.

RAP is an extensible text-based protocol; each request message consist of a method (a sort of command), several headers and message body. We added three new extension methods in the RAP: (1) `GAZE-BIND` to couple the videostreams sent by each participant and define their semantics; (2) `GAZE` method request informs the reflector about the respective user's change of the focus; (3) every client transmitting group video sends `GAZE-GROUP` to establish the site at the reflector and mark the transmitted videostream as group one.

[2] http://mediatools.cs.ucl.ac.uk/nets/mmedia/

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8344
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8344

In the GAZE-BIND request, the following headings are required:

- Front-ssrc: RTP SSRC (media stream identification) of the user's front camera stream
- Side-ssrc: SSRC of the user's side camera stream
- Ip: User's IP address
- Front-port: UDP port where the front camera stream is sent
- Side-port: Same for the side camera
- Audio-ssrc: SSRC of the user's audio stream
- Room: Integer identification of a site where the user is located

The GAZE message contains two mandatory headings: Gaze-who with SSRC of the user, who changed his focus as the argument, and Gaze-at specifying SSRC of the selected user.

Arguments of the GAZE-GROUP message are similar to those of the GAZE-BIND:

- Ip: IP address of the machine sending the group videostream
- Ssrc: SSRC of the group video stream
- Room: Same as in the GAZE-BIND

Besides these extension requests, each personal client also periodically sends the STAT request for a list of participants and their focuses visually highlighting windows of those, who are focused at the respective user. Although gain or loss of someone's focus may be determined from RTP headers of videostreams (which is done in the meantime between two requests), the RAP requests are still needed to find out focus of newly connected clients.

D. Ergonomics

Recent advances in implementation have made the GColl client seemingly as easy to use as original Vic. After short introduction to GColl functionality, testers used to operating Vic seem to operate our client effectively and without difficulties. See part C.2 of the following Section for more details.

E. Used Hardware

For testing of GColl and the user study, we equipped each user with either common wide-screen laptop or a PC with common up to 19 inches sized LCD. Personal videostreams were captured by two Logitech QuickCam Pro 9000 USB webcams at each workstation. Single PC was used for capturing audio and video at each site. The whole group view was recorded by an Elmo PTC-15S tracking camera at both sites. Sound was captured by a ClearOne AccuMic PC echo-canceling microphone at one site and SHURE EasyFlex EZB/C microphones connected to a Gentner AP400 echo-canceling unit at the other.

V. E

We have designed and conducted a quantitative study to evaluate the functionality of GColl. Our aim has been to compare how well could GColl support communication in a complex group task when contrasted to face-to-face and "standard" videoconferencing environments. A "standard" videoconferencing environment in our study consisted of one Elmo PTC-15S tracking camera, a ClearOne AccuMic PC echo-canceling microphone and a projection screen at each site. Audio and video were transmitted by original unmodified Media Tools.

A. General Experimental Design

Overall, there were 90 voluntary participants in the experiment. They were all graduate or undergraduate students of various universities and fields of study. Each received a USB flash disk as a reward for his/her participation. Participants were divided into 15 groups of six members, and each group attended a single experimental session. There were five groups using each communication environment, i.e., talking face-to-face, communicating through the standard videoconferencing environment, and using the GColl environment.

A game called *The Goldminers* was played by the participants at the experimental sessions. This game modeled a real-world problem for the participants and served as a basis for their interaction through the given communication channel. *The Goldminers* is an instantiation of a mixed-motive, social dilemma task, i.e., a situation where apparent gains for each separate individual differ from what would be the best for the whole group. A number of similar tasks has been used in previous research on the topic of computer mediated communication effects (see [16], [3] for examples). The results and the process structure of the game itself can be understood as an indicator of group trust and group coherence, which are usually impaired in computer mediated settings [22]. Each session was videotaped for later analysis after getting a consent from the participants.

The game was facilitated by a small application, which displayed all needed information to participants, calculated the results and logged all actions for their further evaluation. Participants in face-to-face and standard videoconferencing environment were provided with a laptop or a computer with an LCD display for this purpose. In GColl sessions, the Goldminers application was displayed on the same screen as the videoconferencing environment. The game window was designed to be small enough not to interfere with the GColl windows. Based on the task results, participants have received at most two chocolate bars (one per participant in average) as described below in the game rules.

B. Game Description

1) Basic Structure: The Goldminers game is an enhanced version of the well-known Prisoner's Dilemma task. It is an instantiation of a social dilemma, which forces each player as a group member into the decision between the two basic strategies: he can either cooperate with the other group members, which may not be optimal for himself, or try to gain some personal advantage. However, if everyone decides to defect the others, the game score for each single player is worse than if everybody cooperated.

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8344
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8344

2) Rules: In this game, participants represent goldminers and try to mine gold from a river. The river has an attribute called the *gold density* which is set to $30,000 at the beginning of the game.

At the beginning of each round, each participant chooses one of three possible actions: (a) *legal mining*, which gives the participant lower personal profit (current value of *gold density* minus 25 % tax) and causes no harm to the others; (b) *patrolling* the river costs the participant a small amount of gold ($15,000 divided evenly among all participants patrolling that round) while incurring great loss to all illegal miners; and (c) *illegal mining*, which is either worth $50,000, if there was no patrolling in the current round, or causes illegal miner to lose the same amount of money in case a patrol action was chosen. In both cases, *gold density* attribute is decreased by a $1,000 for each illegal miner. Once every three rounds the river partially "cleans itself" thus increasing the attribute by a $1,000.

After all participants choose their action, the round is evaluated and the numbers of actions taken (but not who actually took them) are displayed. The game ends after 15 rounds, or if the *gold density* attribute is ever lower than $1,000. Participants were aware of the exact game ending conditions.

In our version of the game, two ending scenarios were possible: if at least 5 out of 6 participants had more game money than a given threshold ($330,000), participants were awarded with a chocolate bar each; on the other hand, if at least 2 participants did not have enough gold, the group members were given chocolate bars according to their results (first two participants got two bars, the next two got just one bar, and the last two did not receive any chocolate bars). Thus, incentives for cooperative as well as uncooperative play were present.

The threshold value was chosen in such way, that it was just by a few thousands lower than the value attained by the participants if they used only legal mining throughout the whole game. Therefore, the participants had to cooperate quite extensively to reach the threshold at the end of the game – either by playing only the *legal mining* action where after even a slight mistake the scores would be lower than required, or use a coordinated *illegal mining* actions (i.e., that is, the whole group would mine illegally) during some of the rounds to be on the safe side.

Note that there are several substantial differences between the Goldminers and the classical Prisoner's Dilemma: first, any type of communication among the participants is allowed between individual rounds; secondly, there are three possible actions to be taken instead of two; third, the meaning of these actions differ depending on the context (e.g., an *illegal mining* action may be an uncooperative action if everyone else plays *legal mining*, but may be a cooperative action if the whole group decides to mine *illegally*); and finally, the payback for individual actions changes (in a predictable and known way) during the course of the game.

TABLE I

F .

	Mean	Std. deviation
Standard	142 400	12 864
Face-to-face	370 333	96 715
GColl	294 417	82 798

C. Evaluation Indicators

We have evaluated GColl in two dimensions. The first set of indicators focuses on the effect of used environment on the game-play structure and results; second set addresses on the tool usability.

1) Game Based Indicators: Due to the nature of the game, we may understand several aspects of participant behavior during the game as indicators of group trust and/or willingness to cooperate, which are both known to be impaired in standard videoconferencing setting [3] as well as other computer mediated communication (e.g., [22], [15]).

One such aspect are the endgame results of individual participants. The rationale follows from the observation that by cooperating, all members of the group can achieve a high score but if the group members fight among each other, the losses are (in average) larger than the gains. Thus, analyzing whether the mean scores at the end of the game vary for groups using different communication modes can be understood as an indicator of whether the modes tend to inhibit or facilitate group trust. Note that due to the group based nature of the game, a special attention must be paid to distinguishing the effects of communication modes from those created by the interaction inside individual groups.

Mean values and .95CI std. errors of individual endgame scores are shown in the Table I for each communication mode. To analyze the difference among the means, we have performed one-way nested ANOVA [11] over all three communication modes. A normal (i.e., non-nested) ANOVA is inappropriate as the interaction inside each group might have a strong effect on the outcomes on its members. Nested ANOVA test takes this additional variance, if it exists, into account and distinguishes it from the variance arising from difference in communication modes. The test has rejected the null-hypothesis of equality of means with $F(2, 12) = 7.29, p < .05$.

Consequently, this allowed us to perform planned pairwise contrasts between the three communication modes (again using nested ANOVA, instead of using, e.g., a t-test, with the same rationale as above). The Table II summarizes the results of the tests for each pair of modes. These results show that both face-to-face and GColl groups achieved significantly higher mean of gold mined than participants communicating over the standard videoconferencing environment. The difference between means of face-to-face and GColl groups was not significant.

As expected, in both GColl and face-to-face communication modes the participants achieved significantly higher results compared to the standard videoconferencing environment, thus suggesting GColl to be a substantial improvement. On the

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8344
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8344

TABLE II

P

	$F(1,8)$	p
Standard vs. F2F	11.616	$< .05$
Standard vs. GColl	6.654	$< .05$
F2f vs. GColl	1.834	0.213

contrary, the GColl and face-to-face communication modes were not distinguished, which may be attributed to one or both of the following factors: a) the two conditions being similarly efficient for the game used in this evaluation; b) limited sample size of this study.

We have also initially planned to analyze how much the group members tend to keep their word and play the actions they have decided on in the group discussion before each round (all needed information is available in the game logs and video captures of the meetings). However, we have found out that not all groups have accepted a strategy (either for each round separately, or for a longer part of the game), which would be agreed on by all members: in some rounds/groups no strategy has been accepted or even proposed.

Nevertheless, there seems to be a connection between the communication mode used and the overall group behavior (with face-to-face being the "best", that is most of the groups having a strategy and following it; "standard" videoconferencing environment being the "worst" and GColl somewhere in between) but, as this is an indicator on the whole group level, we have too few observations (5 for each communication mode) to make any conclusions. We plan to do an follow-up study to pursue this hypothesis further.

2) Usability Indicators: We have tried to address the possible usability problems of GColl by administering the *Perceived Ease of Use* questionnaire [6], which has been filled out by participants using GColl or "standard" videoconferencing environment; and by semi-structured post-session interviews. In the interviews, no problem with GColl has been repeatedly mentioned by the participants.

To mitigate possible misunderstandings, the questionnaire was translated into the participant's first language. The translation was done by 3 translators, who have reached the final version in several iterations (while following the guidelines in [8]). We have tested the resulting questionnaire during a pilot phase of the experiment, in which 20 participants used the GColl videoconferencing to play the game and filled out the questionnaires afterwards. In a semi-structured group interview, which followed, the participants were asked to discuss any problems they have encountered and correct understanding of individual questions was examined.

Except for question number four[3], where some of the participants understood the term *flexible* in a slightly different way than the rest, no other issues were reported by the test subjects. We have decided to include the fourth question into the final questionnaire without changes and exclude it during the analysis, if needed.

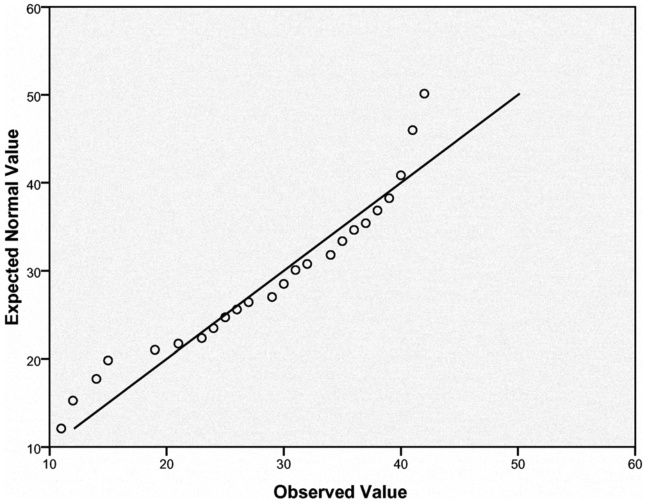

Fig. 3. Normal Q-Q Plot for the Usability Index

The questionnaire is designed as a six item index, each item consisting of one Likert scale (with 7 possible answers). The final value of the index for each participant is calculated as the sum over all questions, thus ranging from 7 (completely unusable) to 42 (without problems). The Cronbach's Alpha is 0.888 for the 6-item scale in our data (0.914 if the question four would be excluded) thus suggesting that all six factors tap the same dimension and the index can be used as intended.

As can be seen at Figure 3, the underlying distribution of our data is most probably not normal – a non-parametric testing needs to be used. The median value was 37 for the GColl participants and 30 in the "standard" videoconferencing condition (the mean scores were 32.2 and 28.0 respectively); Mann-Whitney U test was significant ($p < 0.05$). The results for the index with the question four excluded are very similar.

Our data show that GColl environment felt more usable for the participants, even though they had to, in contrast to the "standard" videoconferencing system, interact with it in a more complex way. Also, the values for GColl are quite high which seems to further confirm the findings from post-session interviews, where no major usability flaws were identified in GColl.

VI. F W

We plan to enhance GColl in the following ways:

A. Eye Tracking

In our current design, users choose other participants by using a mouse or keyboard shortcuts, which might be uncomfortable for some of them. To mitigate this problem, we would like to use the stream from the webcams to estimate direction of user's gaze (e.g., by means of an open-source eye-tracker such as Opengazer, since any hardware eye-tracker would hinder cheapness and flexibility of GColl significantly). We are

[3] The original question is: "I would find *the used system* to be flexible to interact with."

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8344
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8344

attemtping to achieve it without affecting the user's freedom of movement in any significant way (for an illustration of this concept in a different setting see, e.g., [10]). The acquired gaze direction information could then be used for selecting the participant who should be in focus currently.

B. Side Conversations Support

We would like the action of focusing on a remote user to have a further effect. When personal headsets are used instead of whole-group microphone and speakers, we may record speech as well as tailor the reproduced audio individually. Therefore, if a user focuses on a remote participant for a sustained period of time, we are able to increase the volume of the audio stream from that individual participant (as reproduced by the headset) by approximately 15 % of its original level. A similar technique has been used in the GAZE-2 system, where it has been proposed to ease management of side conversations. We are not aware of any rigorous user tests which would support/reject this proposition.

This feature is already implemented in GColl, but has not been evaluated yet. One of the reasons is that we've encountered problems when searching for a suitable headset brand/type: we believe that the headsets should be monaural (as this lowers the feeling of "being cut off" from the colocated participants) and also wireless, if possible. We have tested a number of wireless monaural headsets, but unfortunately up to now, none offered reasonable sound quality.

C. Support of Multiple Sites

We are finishing the support for communication of more than two groups, which will be implemented as an additional logic on the RUM2 packet reflector. The main part, a module which allows composition of several input streams into one output stream, was already implemented in the reflector.

D. Task Support

As we have designed GColl to support group verbal communication, content based task support has been intentionally left out in the initial stages. Even though some basic materials, such as a presentation or a presented document, can be shown on a shared screen, a support for parallel user interaction with the shown documents is missing. At this point, we believe that more elaborate task support could be integrated into the GColl design.

E. A Follow-Up Study

We plan to conduct another medium scale study (100-150 participants) to further evaluate capabilities of GColl. Additionally, there is a case study with real-life users currently underway.

VII. C

We have designed and implemented GColl, a new lightweight videoconferencing environment supporting communication among remote *groups* of people. The GColl design encompasses novel method of conveying mutual gaze and partial gaze awareness using two webcams per videoconference participant.

In order to evaluate feasibility of the GColl design, we have revamped Media Tools Vic and RAT, simple single-person videoconferencing clients to support functionality required by the GColl. New user interface transparently integrates all the video and audio communication into a single tool. Because of need for data processing between senders and receivers, packet reflector is used for multi-point data distribution in GColl videoconferences. We extended the RUM2 modular packet reflector to implement the required functionality; this solution allows us to save significant amount of downstream bandwith at users' workstations, while keeping the clients reasonably simple and their functionality transparent. Both tool and packet reflector, as well as the GColl module, are freely available for download as an open-source project [1].

We have also conducted a user-study with 90 participants to evaluate the functionality and usability of GColl, where no major usability flaws were identified by our group of users. Moreover, GColl has achieved better results than an environment analogical to common commercial systems in both dimensions (i.e., task outcome and environment usability) measured by our evaluation.

When compared to the face-to-face condition, task outcomes of users using GColl were not significantly different from those achieved by participants communicating face-to-face. However, there seems to be a difference in the task process which we want to pursue in follow-up studies.

Acknowledgments

We would like to thank Dr. Stanislav Ježek for his help with the evaluation experiment design and Milan Kabát from Masaryk University for his help with implementation of the GColl prototype. This research has been supported by the CESNET FR 254/2007 project.

R

[1] Gcoll download page at sourceforge.net. http://sourceforge.net/projects/gcoll/.

[2] A. Bezerianos and G. McEwan. Presence disparity in mixed presence collaboration. In *CHI '08*, pages 3285–3290, New York, NY, USA, 2008. ACM.

[3] N. Bos, J. Olson, D. Gergle, G. Olson, and Z. Wright. Effects of four computer-mediated communications channels on trust development. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 135–140. ACM New York, NY, USA, 2002.

[4] N. Bos, J. Olson, N. Nan, N. S. Shami, S. Hoch, and E. Johnston. Collocation blindness in partially distributed groups: is there a downside to being collocated? In *CHI '06*, pages 1313–1321, New York, NY, USA, 2006. ACM.

[5] M. Chen. Leveraging the asymmetric sensitivity of eye contact for videoconference. In *CHI '02*, pages 49–56, New York, NY, USA, 2002. ACM.

[6] F. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 1989.

[7] J. Denemark, P. Holub, and E. Hladká. RAP – reflector administration protocol. Technical Report 9/2003, CESNET, 2003.

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8344
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8344

[8] J. Harkness. Questionnaire translation. *Cross-cultural survey methods*, pages 35–56, 2003.

[9] E. Hladká, P. Holub, and J. Denemark. User empowered virtual multicast for multimedia communication. In *ICN'04*, pages 338–343, Guadeloupe, 2004.

[10] A. Hyrskykari, P. Majaranta, and K. Räihä. Proactive response to eye movements. In *Human-computer Interaction: INTERACT'03; IFIP TC13 International Conference on Human-Computer Interaction, 1st-5th September 2003, Zurich, Switzerland*, page 129. Ios Pr Inc, 2003.

[11] R. Mason, R. Gunst, and J. Hess. *Statistical Design and Analysis of Experiments, 2nd Edition*. 2003.

[12] A. Monk and C. Gale. A Look Is Worth a Thousand Words: Full Gaze Awareness in Video-Mediated Conversation. *Discourse Processes*, 33(3):257–278, 2002.

[13] D. T. Nguyen and J. Canny. Multiview: improving trust in group video conferencing through spatial faithfulness. In *CHI '07*, pages 1465–1474, New York, NY, USA, 2007. ACM.

[14] K.-I. Okada, F. Maeda, Y. Ichikawaa, and Y. Matsushita. Multiparty videoconferencing at virtual social distance: Majic design. In *CSCW '94*, pages 385–393, New York, NY, USA, 1994. ACM.

[15] G. Olson and J. Olson. Distance matters. *Human-computer interaction*, 15(2):139–178, 2000.

[16] E. Rocco. Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 496–502. ACM Press/Addison-Wesley Publishing Co. New York, NY, USA, 1998.

[17] A. Sellen, B. Buxton, and J. Arnott. Using spatial cues to improve videoconferencing. In *CHI '92*, pages 651–652, New York, NY, USA, 1992. ACM.

[18] P. Slovák, P. Troubil, and P. Holub. Gcoll: A flexible videoconferencing system for group-to-group interaction. In *INTERACT 09*, 2009. To appear.

[19] D. Tutt, J. Hindmarsh, M. Shaukat, and M. Fraser. The Distributed Work of Local Action: Interaction amongst virtually collocated research teams. In *Proc. ECSCW*, pages 199–218, 2007.

[20] R. Vertegaal, G. van der Veer, and H. Vons. Effects of Gaze on Multiparty Mediated Communication. *Graphics Interface*, pages 95–102, 2000.

[21] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In *CHI '03*, pages 521–528, New York, NY, USA, 2003. ACM.

[22] J. Walther and U. Bunz. The rules of virtual groups: Trust, liking, and performance in computer-mediated communication. *The Journal of Communication*, 55(4):828–846, 2005.

Digital Object Identifier: 10.4108/ICST.COLLABORATECOM2009.8344
http://dx.doi.org/10.4108/ICST.COLLABORATECOM2009.8344