

CollaBoard: A Novel Interactive Electronic Whiteboard for Remote Collaboration with People on Content

Andreas Kunz, Thomas Nescher, Martin Kuechler

Department of Mechanical and Process Engineering
ETH Zurich

Tannenstrasse 3, 8092 Zurich, Switzerland
{kunz, nescher, kuechler}@iwf.mavt.ethz.ch

Abstract—In this paper, we present a device called ‘CollaBoard’. It was developed in the context of ongoing CSCW research efforts in developing groupware that mediates remote collaboration processes. CollaBoard combines video- and data-conferencing by overlaying a life-sized video showing the entire upper body of remote people in front of the displayed shared content. By doing so, CollaBoard shows pose, gaze, and gestures of remote partners, preserves the meaning of users’ deictic gestures when pointing at displayed shared artifacts, and keeps shared artifacts editable at both conference sites. For this, a new whiteboard software is also introduced, which allows a real-time synchronization of the generated artifacts. Finally, the functionality of two interconnected CollaBoard prototypes was verified in a usability assessment.

Keywords - remote collaboration; shared workspaces; computer supported collaborative work; digital whiteboard; mixed presence groupware;

I. INTRODUCTION

Remote collaboration by using video- and data-conferencing systems becomes increasingly popular. As a result, software clients that were initially designed for video-conferencing nowadays also allow sharing a common workspace between conferees for collaboration purposes (data-conferencing). Working together, however, is much more than just seeing each other and sharing a view of the content that is elaborated (images, drawings, slideshows, etc). In collaboration processes, being co-located or remote, the focus of attention is mainly on the shared artifacts (or more general: task-centered) [4], and collaborators use deictic gestures for referring to shared artifacts. The importance of body languages such as (deictic) gestures, pose, (eye) gaze, and facial expressions of remote collaborators were already identified earlier [5, 6, 7, 8, 9]. However, commercially available conferencing systems do not allow e.g. a natural use of deictic gestures during combined video- and data-conferencing; since they display video and shared workspace in separate application windows, thus breaking the spatial connection of deictic gestures to the content in the shared workspace.

The development and implementation of conferencing systems that enable rich use of body language in remote collaboration processes are subject to ongoing CSCW research. The paper in hand fits in this research and presents CollaBoard, a conferencing system that gives distant collaborators the sensation of being virtually next to each other, and that allows collaborators using (deictic) gestures, gaze, and pose for communication (see Figure 1).

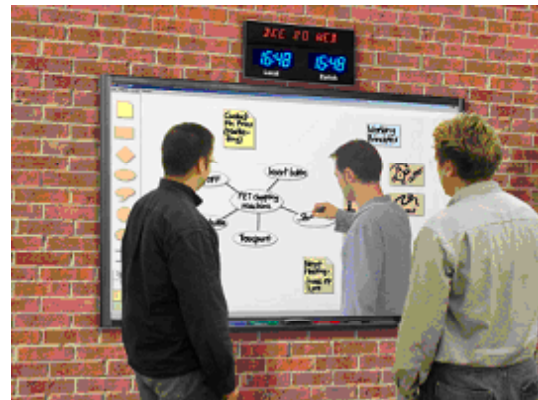


Figure 1. The CollaBoard idea.

II. RELATED WORK

An early work in this particular field of CSCW research was e.g. ClearBoard [5]. A recently presented, very elaborated conferencing system that supports collaborator’s natural use of gestures is VideoArms [10, 11]. The system acquires people interacting on shared workspaces (large interactive displays) by means of a camera that is on-axis with the display device. Therefore, the context of hand gestures is preserved, e.g. deictic gestures pointing out a shared artifact can be correctly interpreted by the remote collaborator. VideoArms is limited due to the deployed segmentation algorithm. While the system transmits a video showing hands and arms, it fails to transmit the collaborator’s upper body which would mediate additional consequential and inconsequential communication such as

pose and gaze that is not presented by the arms. Alternative segmentation algorithms are available [12], but entail time-consuming calculations, which lead to stuttering live video embodiments.

E-Chalk enhanced with SIOX [3] provides a very pleasing live video embodiment by using an elaborated segmentation technique that includes information from a depth-sensing camera. In order to mitigate any occlusion, the developers of the system chose to display the live video embodiment in a slightly translucent way. Note, however, that E-Chalk was developed to support distant learning, and not remote collaboration sessions. As a consequence, the system only provides a unidirectional live video embodiment (showing the teacher), and artifacts are not meant to be editable at both conferencing sites.

III. CONTRIBUTION

To overcome the limitations of VideoArms and E-Chalk with SIOX, we designed and implemented a device called ‘CollaBoard’. It provides video- and data-conferencing with live videos showing the full upper bodies of remote collaborators. The artifacts in the shared workspace are editable for all collaborators. By displaying the videos atop the shared workspace, the meaning of collaborators’ (deictic) gestures preserve their meaning relative to the shared artifacts.

With the CollaBoard, video and content are acquired and transmitted separately. Content (i.e. the artifacts in the shared workspace) is transmitted as application data to keep it editable at both conference sites. The collaborators are acquired by a video camera that is positioned opposite to the screen. The acquired video stream is sent to the computer, processed, and transferred to the remote conference site. Video processing includes segmenting the user (foreground) from his or her background. The segmentation is necessary for subsequent video overlay atop the shared content at the remote conference site.

Note that the background of the acquired video images consists of the shared workspace, therefore being highly inhomogeneous and also dynamic, which complicates the segmentation. To avoid this problem, we blank the shared workspace to the camera, while keeping it fully visible to the collaborators. A popular method to do so is the use of linearly polarized light and matching filters [13, 20]. We also use this method for our CollaBoard setup.

For the segmentation, we use an algorithm based on the illumination invariant method [14]. Assuming a now static background, previously captured sequences of the blanked shared workspace without any user in front of it are compared to the actual image. This is done by using a statistic criterion that measures the colinearity of the actual color and the expected background in the color space. Pixels identified as background are chroma-keyed by changing their colors to green. The resulting image is compressed as part of the video stream, and transferred to the remote conferencing site.

At the remote conferencing site, the received chroma-keyed video stream is overlaid atop the shared content by

rendering green pixels in the video as transparent, and displaying the video in front of the content.

IV. HARDWARE IMPLEMENTATION

For our research, we designed and realized two identical CollaBoard hardware setups. Each setup consists of a large display, interaction module, camera and filter, lighting bar, an audio system, and a computer with network access.

For each of our two CollaBoard prototypes, we use a 65” widescreen liquid crystal display (LCD). LCDs emit linearly polarized light, a feature which we use for blanking the displayed content to the camera. Since the camera is also equipped with a linear polarization filter (which is rotated by 90° with regard to the polarization of the incoming light from the LC-screen), it cannot capture the content of the whiteboard anymore, but sees a very dark gray instead. However, the user standing in front of the whiteboard, randomly reflects incoming light and thus is still visible to the camera.

To use the display as an electronic whiteboard, it must be made interactive. We use a commercially available interaction module, which allows touch- and pen-based on-screen interaction. A pen tray at the lower display border provides pens in four colors, as well as an eraser. The module’s protective glass panel was replaced by another one that preserves the polarization of transmitted light.

The interactive overlay is mounted on the frame of the LC-screen and thus the overlay’s glass pane is in a certain distance to the LC-matrix. This results in a noticeable distance between the interaction plane and the image plane. The so-called parallax distortion is the error between the perceived position of an object and its effective position. It results from the distance between the image plane (LC-matrix) and the interaction plane (glass of the DVIT overlay) [18]. In addition, the parallax distortion has a significant impact on the interaction’s user friendliness, since the interaction’s precision is more and more distorted with increasing distance between image plane and interaction plane. Small objects (targets) cannot be hit easily and the user has to continuously correct the position during his interaction.

In order to reduce the gap, the display was completely disassembled. Within the reconstruction, all spacers were removed and now the LC-matrix is only held in place with the DVIT-overlay. Thus, the distance between image plane and interaction plane was reduced from 15 mm to 7 mm.

A preliminary user study showed that an offset between the image plane of the display device and the interaction plane of the input device affects the input precision. A total of 50 participants (undergraduate and graduate students from our university) took part in the study. The participants were divided into 19 groups of two, and 3 groups of four people. The participants of each group were engaged in a remote collaboration to solve a puzzle task [19]. The task required the remote participants to collaborate intensively in order to successfully solve it. After the task completion, the participants were asked to give a short written feedback, and to specify things which disturbed them, if any. The gathered data clearly shows how the input performance can be

increased by reducing the said offset. Figure 2 shows some results of the multi-directional tap test, resulting from the application of the ISO 9241-9 standard.

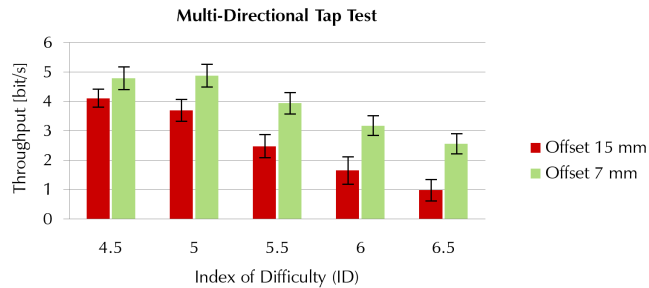


Figure 2. Measurements from the multi-directional tap test.

V. SOFTWARE IMPLEMENTATION

Typical video conferencing systems transfer the video image and the content of the whiteboard time-sequentially, or they display the video and the whiteboard's content on different screens. Thus, the video and the whiteboard's content (the workspace) are decoupled. This means that the video and the content do not have to be perfectly synchronized all the time. Thus, today's whiteboard applications transfer content to the remote partner only at certain intervals. A continuous, real-time synchronization is complex to implement and requires a higher bandwidth. For instance, in most existing whiteboard applications, a stroke that has been drawn with a pen is only transmitted after the pen is removed again from the whiteboard's surface.

In the CollaBoard case, the video is part of the whiteboard content and therefore the whiteboard content must be synchronized in real-time between the remote peers. For example, if a local peer draws a line on the whiteboard, the video at the remote location shows him moving a pen on the whiteboard. In order to give the remote peer the sensation that a line is being drawn, the line must show up in real time while it is being drawn and in sync with the video. We call this feature real time strokes.

In this context, a stroke is regarded as one interaction, no matter if it is a drawing or erasing interaction. Hence, a stroke is one movement of a tool on the drawing layer from the time when the tool becomes active until it is deactivated or removed again.

A. Audio- and video connection

For the audio link between the CollaBoards, Skype audio-conferencing software is used [2]. The video connection is established by running a customized version of the open source software ConferenceXP [15]. Software customization includes the integration of self-programmed modules for video processing capabilities (video segmentation) such as lens distortion correction, segmentation, and video overlay.

In principle, the idea of image segmentation is to decide, whether a pixel in the image belongs to the foreground or to the background. For doing so, each pixel of the acquired image is compared to the one of a reference image. If the

"difference" between these two pixels is below a certain threshold, it is interpreted as noise and the pixel is assigned to the background. Thus, the calculation of this "difference" is crucial. Several statistic models and correlation functions exist, which can be used for the segmentation.

In a next step, outliers need to be eliminated. Outliers are single pixels and smaller areas that are classified significantly different from neighboring pixels and areas. If for example a threshold of x pixels is defined, all blobs will be removed, whose areas are smaller than x pixels. This process is also known as the application of morphological filters.

The same procedure cannot only be used between neighboring pixels within an image, but also between subsequent images within a video sequence. If for example a pixel is classified to belong to the background, it also should be a background pixel in the next image as long as there was no abrupt change. This prevents rapid changes of pixels between foreground and background, which would be visible in a video sequence as occurring noise, which is an unwanted effect.

The algorithm described in the above was implemented under WindowsXP and uses the multimedia framework DirectShow. Figure 3 shows an ideally segmented image, in which the background was replaced by green.

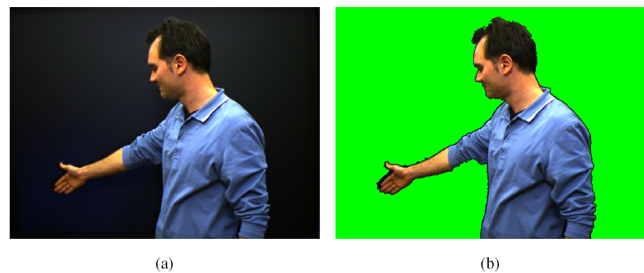


Figure 3. Segmentation (a) Original image, (b) Segmented image with a modified green background.

B. Whiteboard Software

Many client-server or peer-to-peer based whiteboard applications provide synchronization of strokes across the network. But synchronization is done at the stroke level so that only finished strokes or user interactions have to be transmitted. Although this makes the synchronization significantly easier, it is not helpful for the CollaBoard and also introduces several problems. What happens for instance, if two peers are drawing something at the same time at the same place? Both will not notice that their partner is drawing as well, because the transmission takes place only after they will have removed the pen. Such a stroke-based synchronization can be regarded as a time-multiplexed synchronization method. Strokes are sent to the server and then distributed to all the clients. The last stroke arriving at the server stays on top of previous strokes. In order to support real time strokes, there are two possibilities. One way is to artificially cut strokes into smaller strokes, i.e. during a user interaction. Now, a movement is artificially split up into several smaller movements (the pen is virtually

removed and set again). The shorter the time interval between the invisible splits is, the more strokes are created by one user interaction, and the more transmissions take place. Therefore, the interaction is transmitted in almost real-time, if the interval is small enough. The problem with this approach is that it generates a lot of overhead. The stroke database is filled rapidly, and erasing is much more complicated, as searches in the stroke database become more time-consuming. Furthermore, the stroke's smoothness is lost.

We developed a new digital whiteboard application, which offers a plane white area for drawing. Our software is based on the InkCanvas class provided by Microsoft's .NET environment [16]. The physical pens from the pen tray can be used to draw on the drawing layer and to choose a color. The eraser can be activated as well by using the tray. Images can be loaded onto the background and are automatically sent to the server and distributed to all other connected clients. At the bottom of the window, there are two buttons for switching pages. Clicking the page forward or backward button results in a page change request to the server. The server stores the current page, switches to the new one and sends an update to all clients. Figure 4 shows a screenshot of our whiteboard application.

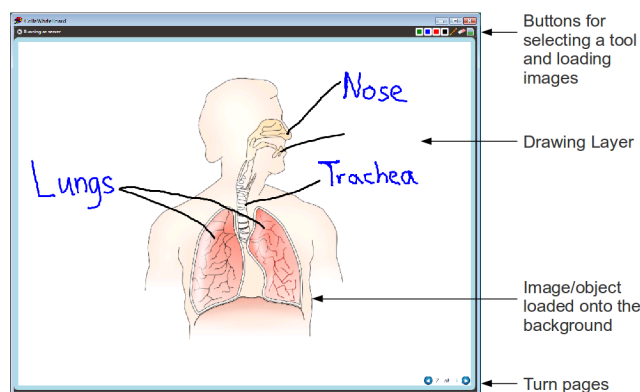


Figure 4. Screenshot of the whiteboard application.

For our whiteboard application, we chose a better way to support real time strokes than suggested above. The transmission of strokes is split up into a two-phase process - much like InkCanvas does for visualization. During a user interaction, the coordinates of the movement are recorded by a real-time transmission module (as InkCanvas does for dynamic rendering) and sent to the remote peer (the update rate of the transmitted coordinates can be reduced for slow network connections). The peer receives the points and renders them with a dynamic renderer on a drawing layer in the background. This means that the dynamically rendered content of the peer remains in the background (but in front of background objects/images), while the local content stays in the foreground. Therefore, a drawing user is not disturbed by his partner while sketching. As soon as the user interaction is finished, the stroke is generated and sent to the server as above. If a client receives a remote stroke from the server, the according dynamically rendered content (of the remote

peer) is replaced by the stroke in the foreground. This replacement is almost invisible to the user, except for the fact that the stroke comes to the foreground. Hence, our software allows two or more users to work on visual content from different locations at the same time.

In summary, the whiteboard application handles storage, display and the transmission of strokes in a two-phase process. Stroke creation, storage and visualization are done by InkCanvas. A stroke is created after a drawing interaction is finished and the tool is inactive again. Therefore, once a tool starts moving on the drawing layer, the coordinates of the movement are stored by InkCanvas. When the movement is finished, a curve is fitted to the coordinates and the curve along with its points is stored as a stroke in the stroke database. Visualization works similarly as a two-phase process. The dynamic renderer of the InkCanvas framework continuously draws a line while a tool moves on the drawing layer. Hence, it seems as if the ink "flows" out of the pen's tip. Once the stroke is stored, the dynamically rendered stroke is replaced by the static renderer with a smooth curve.

VI. USER STUDY

We conducted a user study in order to analyze the performance of the CollaBoard system.

The user study is designed to be similar to a real life scenario, in which two remote groups have to work together on a common problem. We believe that the new features of the CollaBoard offer a better and more natural way of collaboration among remote partners and therefore we compare it to other systems (see TABLE I.).

A. Experimental Design and Apparatus

We decided to conduct a user study with three different conditions. One condition is the CollaBoard system, another one is a condition that is similar to standard video conference software. Finally, we also evaluated a control condition, in which the partners or users were working together at the same location and on the same whiteboard.

The hardware used for all conditions of the study was the CollaBoard screen with the tool tray and the interactive frame from SMART Technologies. Users could use the tool tray to select one of four colors or an eraser. Only one tool can be used at a time. The CollaWhiteBoard software was used for all conditions, but with different configurations.

1) Collocated Condition (CO)

In the collocated condition, both groups, which are usually working at remote locations, work together at the same CollaBoard in the same room. They share one whiteboard and therefore cannot sketch at the same time. For sketching, the CollaWhiteBoard software is used. The collocated condition simulates the optimal case, in which no remote collaboration is required and thus serves as a control condition for the user study.

2) Separate Video Condition (SV)

In the separate video condition, a setup is simulated which is equal to traditional video conference software. Two groups are working in two different rooms with two CollaBoards. The CollaWhiteBoard software is used, but with the real time strokes feature turned off, i.e. if a user

draws a stroke, the remote partner will see the stroke only after the local user removes the pen from the CollaBoard. In addition, there is an audio and video link with Skype [2]. The video of the remote user is presented on an extra 19" screen, which is placed on the left side of the CollaBoard. For the video link, a Logitech Pro9000 webcam sits on top of this extra screen, which records the upper body part of the CollaBoard user.

3) CollaBoard Condition (CB)

In the CollaBoard condition, all CollaBoard features are used as described above. Two groups are working in two different rooms with two CollaBoards. Skype in contrast to the SV condition is only used for audio. The video is recorded with the CollaBoard software and presented on top of the whiteboard content. The CollaWhiteBoard software has the real time strokes feature turned on. The external 19" screen and the webcam are not used, but the CollaBoard camera is active.

TABLE I. OVERVIEW OF USER STUDY CONDITIONS AND DIFFERENCES.

Condition	Location	Video	Sketching	Audio
Collocated (CO)	both groups local in same room, 1 whiteboard	no video	CollaBoard whiteboard software, only one user at a time	
Separate Video (SV)	2 rooms/groups	video on separate screen, webcam in front of user	CollaBoard whiteboard software, without real time strokes	Skype audio link
Colla-Board (CB)	2 rooms/groups	video on content, CollaBoard camera at the back of the user	CollaBoard whiteboard software with real time strokes	Skype audio link

B. Participants

We recruited fourteen subjects (12 male and 2 female) for the user study (median age 22.5 years). For each study, a group of two subjects had to work together on the study tasks. Hence, we had seven groups for the study. Four groups were students and three groups were staff members of the mechanical engineering department. All subjects knew each other, but have not been working together on a digital whiteboard before. Staff members were familiar with digital whiteboards and basic conferencing tools (as in the SV condition). Students were given a general introduction and training phase of about two hours. No subject worked with the CollaBoard system before.

Beside the two persons working at the whiteboard, more subjects were present as local audience. The local audience consisted of students as well as staff members, who were not actively working at the whiteboard. For each study, between two and eight subjects were present as local audience.

C. Task

We designed the study task to be similar to a real life scenario. The task should enforce the users to work collaboratively on a common problem and also provoke them to communicate a lot. Having a lot of interaction between the two users is crucial for evaluating the major improvements of the CollaBoard over classical conference systems. This is motivated by the media richness theory [17].

In our study, we asked the participants to design a floor plan for a house. Both participants got the same task, but with different requirements. Task A for instance is to draw the floor plan of a new mansion in Zermatt. Participant 1 has the requirement that the main entrance should be on the south side. Participant 2 instead has the requirement that there is an entrance hall in the center of the house. Hence, the two remote partners have asymmetric information about the house they have to plan. Each participant has five requirements about the house which are different from his partner's requirements. They do not have sufficient information to design the house on their own. Both sets of requirements complement each other to a common solution. The uncertainty about the remote partner's requirements helps to enforce close collaboration.

The requirements are sorted from more general ones first to specific ones at the end. Therefore, the participants were asked to go through the given requirements stepwise.

We created three different floor plan design tasks for the three different conditions.

D. Experimental Setup

We prepared two rooms with equal setups as shown in Figure 5. Chairs were prepared for the local audience and placed to the left and right of the CollaBoard camera. For the SV condition, a 19" display was placed to the left of the CollaBoard with a webcam on top. The CollaBoard user himself was able to move freely in the area between the screens and the local audience. The microphone was placed next the CollaBoard tool tray and the loudspeakers were placed below the CollaBoard.

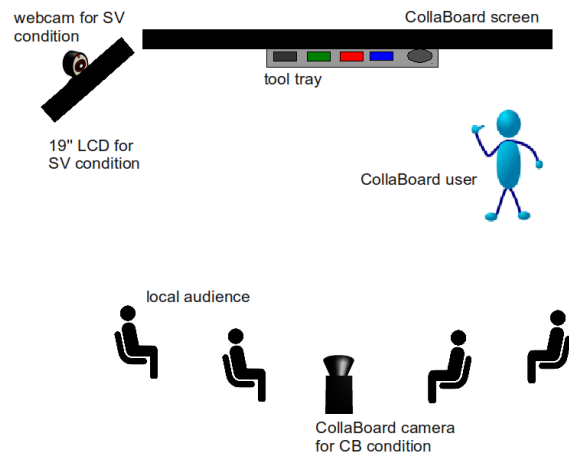


Figure 5. Setup of the user study for one room.

TABLE II. AVERAGE SCORES AND STANDARD DEVIATION (SD) FOR ALL 13 QUESTIONS OF THE USER STUDY QUESTIONNAIRE. AT THE END OF THE TABLE, THE MEAN OVER ALL QUESTIONS IS GIVEN.

Question	CO		SV		CB	
	Mean	SD	Mean	SD	Mean	SD
1. I could easily make myself well understood	4.36	0.74	3.64	0.84	3.79	0.89
2. I could easily tell what my partner was pointing at	4.43	1.09	2.57	1.09	3.29	0.73
3. I could easily tell what my partner was referring to (verbally or by pointing gestures)	4.43	0.94	3.00	0.96	3.21	0.89
4. I found it easy to interrupt my partner when I saw him make an error	4.29	1.14	3.29	0.91	3.43	1.09
5. I could easily tell what my partner was looking at	4.43	1.09	2.14	0.86	2.86	0.95
6. My remote partner and I continuously communicated with each other	4.64	0.50	3.93	1.14	4.00	0.96
7. I was never confused	4.21	1.12	3.21	0.80	2.86	0.95
8. We never spoke at the same time	3.36	1.15	3.21	1.05	3.07	1.07
9. I often looked at my partner	4.00	1.18	2.36	1.08	2.86	1.35
10. I exactly knew when it was my turn to speak	3.93	1.00	3.14	0.95	3.07	1.00
11. I could easily tell if my partner was listening carefully to what I said	4.21	0.97	2.79	1.37	3.00	1.18
12. When I looked at my partner, I could always clearly see his or her face	4.50	1.16	2.86	1.41	1.86	0.95
13. I am happy with the elaborated floor plan	4.36	0.93	3.50	0.94	3.79	1.05
Mean over all questions in questionnaire	4.24	1.03	3.05	1.12	3.16	1.11
Mean over all questions from questionnaire without question 7 and 12	4.22	1.02	3.05	1.12	3.31	1.06

For all three conditions, the two CollaBoard users were working on the same shared content. For the CO condition, both users were working in the same room on one CollaBoard.

E. Procedure

At the beginning of the study, two supervisors informed the participants about the task they were asked to solve. They were told that they have to solve three similar tasks on three different conditions. One condition is both working together locally and the other conditions are that they work together on the same problem in two different rooms with conference software. The local audience, which was formed by other students, was split into two groups of equal size as well.

We kept the order of the conditions the same for all studies, but we randomized the order of the three tasks according to a Latin square scheme. The ordering of the conditions was CO, SV, and finally CB. We did not have enough participants to rotate the order of conditions as well.

Before each study on one condition, we quickly explained the features of the current condition as outlined above. Then, we handed out the task sheets to the CollaBoard users and to the local audience. The local audience could participate in solving the task by giving hints to the CollaBoard user in the same room. But they were asked to remain seated throughout the study and could not join the CollaBoard user at the whiteboard. The participants had two minutes to read the task and then the CollaBoard users had to start solving the task.

The participants were told to solve the task quickly. The task was defined to be solved if all requirements for the floor plan were met by both groups. For the SV and CB condition, the participants were also told that they can work (sketch) simultaneously on a common solution.

In the CO condition, both groups and the whole local audience stayed in the same room. The participants from both groups were not allowed to exchange their task sheets. After finishing the first task, the second group was asked to move to another room.

After all three tasks were solved under all conditions, the users working at the CollaBoard were asked to fill out a questionnaire with equal questions for the three setups. Figure 6. shows a typical situation on the CB condition.

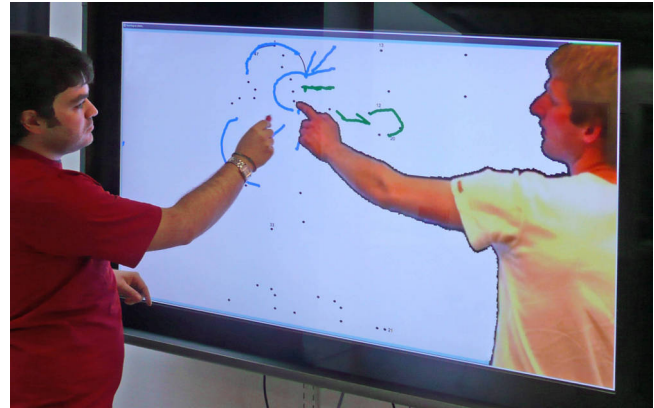


Figure 6. Typical situation in the CB condition

VII. RESULTS

All groups managed to design correct floor plans and solved all tasks on every condition. No group had to be interrupted because they took too long. As suggested by the supervisors, all participants solved the task in a stepwise manner by starting with the first requirements first.

A. Expectations

Primarily we expected the CO condition to perform best. We hoped that the CB condition performs better than the SV condition and that there is a significant difference between the SV and CB condition.

B. Questionnaire Results

The questionnaire the participants had to fill in after finishing all three tasks contained 13 questions for evaluating the usability. Each question had to be answered three times,

once for each condition. The scale of the questionnaire is a Likert-type scale from 1 (“strongly disagree”) to 5 (“strongly agree”). The questions were partially taken from [1] and modified to fit the CollaBoard user study. TABLE II. presents the questions, the mean response per question and setup, and the standard deviation. The questionnaire was designed to allow for a summative analysis.

C. Supervisor Measures Results

During the study, additional measures were recorded by user study supervisors. There were two supervisors present, one for each room/participant. The results are presented in TABLE III. below.

1) Total Task Completion Time

The completion time is defined from the moment when the participants finish reading the task for the first time and agree that they were both ready to start, until the moment when they both agree that all their requirements are met and that the task is solved therefore.

2) Number of Turn Takings (non-verbal)

Whenever the local user starts acting again, i.e. making a pointing gesture or drawing on the screen, the number of turn-takings is increased by one.

3) Number of Pointing Gestures

A pointing gesture is whenever the local user refers to one object on the screen by pointing at it. So, if the user refers to two objects or locations (i.e. “this object here and this one there”), even with one movement of his hand on the screen, this is counted as two pointing gestures. Drawing actions are not counted.

TABLE III. MEAN VALUES AND STANDARD DEVIATION (SD) OF SUPERVISOR MEASURES.

Measure	CO		SV		CB	
	Mean	SD	Mean	SD	Mean	SD
Completion time in minutes	6.71	1.20	7.14	2.57	5.86	1.88
Number of turn takings	5.14	2.44	6.50	2.21	6.57	2.53
Number of pointing gestures	5.29	2.43	3.29	3.69	4.00	3.21

VIII. DISCUSSION

As expected, the CO condition outperforms the SV and CB conditions. But nevertheless, even the CO condition did not achieve an average score of 5 (strongly agree) on all questions. According to different statements of the participants, this is due to social issues between the two partners working at the whiteboard. As mentioned above, the participants have not been working together at a digital whiteboard before. In several cases in the CO condition, one user strongly dominated at the whiteboard, i.e. one user was drawing all the time while the other one just read out the requirements on his task sheet and did not participate actively in drawing the floor plan. Such a strong domination

by one participant could not be observed during the SV or CB condition.

As also expected, the SV condition performs significantly worse than the CO condition for all questions and completion time.

Our assumption that the CollaBoard performs better than the SV condition cannot be proven with a summative analysis over all questions, which is probably due to the small amount of users we evaluated so far. On average, the CB condition performs better than the SV condition, but the variance between the data is quite high (see TABLE II.). On four questions, the CB condition is rated worse than the SV condition. As expected, especially for question 7 and 12, the CollaBoard is rated significantly worse than the SV condition. By design, question 12 cannot be rated higher than in the SV condition. As in the CB case, the video is on top of the content. The user working at the whiteboard usually does not see the face of his opponent directly (face to face), but from the side. He will see gestures very well instead. The remote user's face is clearly visible to everyone if he looks at his local audience. In the SV condition, the local user can see his opponent clearly but not in relation to the content and only if he decides to look at the webcam. Therefore, a user has to stop working to look at his partner in the SV setup.

Similar to question 7, the video on content feature confused many users. Most participants have been using Skype video conferencing at home before, but they all have not seen or used any video on content conferencing solution before. The confusion usually came from the fact that they were recorded from the back and were not aware of the fact that their body is inserted into their remote partner's content, even though they could see their partner on the screen. Therefore, some participants kept blocking the screen so that the remote partner could not work.

Hence, for a summative analysis, we decided to ignore question 7 and 12 (see Figure 7.). In this case, the CollaBoard performs better than the SV condition.

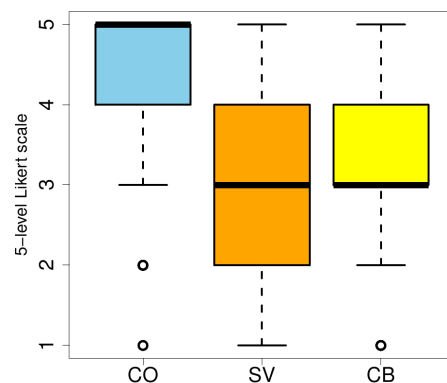


Figure 7. Summative analysis of all questions except question 7 and 12. The boxplot shows the median value. The lower and the upper end of the box are the 25th and the 75th percentile. The whiskers extend to 1.5 times the IQR. Points mark outliers.

Because of too few participants, we could not rotate the order of the conditions. Hence, we had a training effect across the three conditions. This can be seen by looking at

the completion time (see TABLE III.). The CB condition performs best, but it was also the last condition to be evaluated. But here, the SV condition is clearly worse than the CB condition, which implies that the CollaBoard is better for solving the task more efficiently.

So far, it is hard to draw any conclusions from the number of turn takings or number of pointing gestures because the variance in the data is too high.

IX. CONCLUSION AND OUTLOOK

We presented the CollaBoard - a novel whiteboard and conference system for remote collaboration. By employing modern techniques like foreground segmentation, overlaying people onto the content and real-time synchronization of user generated artifacts, the CollaBoard provides a better user experience compared to traditional remote conference systems with whiteboards.

In a first usability evaluation, we compared the CollaBoard to traditional systems. We showed that the CollaBoard performs well and has a great potential.

For a detailed comparison with traditional conference solutions, a more elaborate user study is planned with more participants.

Further research should analyze the experience of the local audience and their impact on the usability. Furthermore, a system should be evaluated that combines the CollaBoard with a separate video condition.

ACKNOWLEDGMENT

This work was done within the Eureka project $\Sigma!$ 4066. We want to thank all people who contributed to this work.

REFERENCES

- [1] J. Hauber, H. Regenbrecht, M. Billingham and A. Cockburn, "Spatiality in videoconferencing: trade-offs between efficiency and social presence", Proc. of the 2006 20th Anniversary Conference on Computer Supported Cooperative (CSCW '06). ACM, New York, pp. 413-422. doi:10.1145/1180875.1180937
- [2] Skype, URL: <http://www.skype.com> (accessed May 9, 2010).
- [3] G. Friedland, "Adaptive Audio and Video Processing for Electronic Chalkboard Lectures," PhD Thesis, Department of Computer Science, Freie Universität Berlin. 2006.
- [4] W. Gaver, A. Sellen, C. Heath, P. Luff, "One is Not Enough: Multiple Views in a Media Space," in Proc. CHI 1993, ACM Press; 1993, pp.335-341.
- [5] H. Ishii, K. Arita, M. Kobayashi, "Towards Seamless Collaboration Media – From TeamWorkStation to ClearBoard," in NTT Review Vol. 5, 1, 1993, pp. 24-29.
- [6] D. Kirk, T. Rodden, D. Stanton Fraser, "Turn It This Way: Grounding Collaborative Action with Remote Gestures," in Proc. CHI 2007, ACM Press, 2007, pp. 1039-1048.
- [7] D. Kirk, D. Stanton Fraser, "Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks," in Proc. CHI 2006, ACM Press, 2006, pp. 1191-1200.
- [8] M. Louwerse, A. Bangerter, "Focusing Attention with Deictic Gestures and Linguistic Expressions," in Proc. CogSci 2005, Lawrence Erlbaum Associates, 2005, pp. 1331-1336.
- [9] J. Tang, "Findings from Observational Studies of Collaborative Work," in Int. Journal of Man Machine Studies 34, 1991, pp. 143-160.
- [10] A. Tang, C. Neustaedter, S. Greenberg, S. "VideoArms: Embodiments for Mixed Presence Groupware," in Proc. HCI 2006, ACM Press, 2006, pp. 85-102.
- [11] A. Tang, C. Neustaedter, S. Greenberg, "VideoArms: Supporting Remote Embodiment in Groupware," in Video Proc. CSCW 2004, ACM Press, 2004.
- [12] F. Coldefy, S. Louis-dit-Picard, "Remote Gesture Visualization For Efficient Distant Collaboration Using Collocated Shared Interfaces," Proc. IASTED HCI 2007, ACTA Press, 2007, pp. 37-42.
- [13] J. Tang, S. Minneman, "VideoDraw: A Video Interface for Collaborative Drawing,". In Proc. CHI 1990, ACM Press, 1990, pp. 313-320.
- [14] R. Mester, T. Aach, L. Dümbgen, "Illumination-Invariant Change Detection Using a Statistical Colineary Criterion," in. Proc. DAGM 2001, Springer, 2001, pp. 170-177.
- [15] Microsoft Research. ConferenceXP Project. URL <http://research.microsoft.com/conferencexp> (accessed May 9, 2010).
- [16] Microsoft InkCanvas Class, .NET framework 3.5. URL <http://msdn.microsoft.com/en-us/library/system.windows.controls.inkcanvas.aspx> (accessed May 9, 2010).
- [17] R.L. Daft, R.H. Lengel, "Information richness: a new approach to managerial behavior and organizational design," Cummings, L.L. & Staw, B.M. (Eds.), Research in organizational behavior 6, 1984, JAI Press, Homewood IL, 1984, pp. 191-233.
- [18] DVIT – Digital Vision Touch Technology – White paper. Available: http://smarttech.de/dvit/DVIT_white_paper.pdf, accessed 27. Sep. 2008
- [19] M. Billingham, D. Belcher, A. Gupta, K. Kiyokawa, "Communication Behaviors in Co-located Collaborative AR Interfaces," in International Journal of Human-Computer Interaction 16, 3, 2003, pp. 395-423.
- [20] T. Piazza, M. Fjeld, "Ortholumen: Using Light for Direct Tabletop Input." in Proc. Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'07), 2007, pp. 193-196.