Guided Design and Evaluation of Distributed, Collaborative 3D Interaction in Projection Based Virtual Environments

Gernot Goebbels[1], Vali Lalioti[2], Thomas Mack[1]

[1] German National Research Center for Information Technology
Department of Virtual Environments
[2] Computer Science Department
University of Pretoria, South Africa

ABSTRACT

In this paper we present a framework for the design and evaluation of distributed, collaborative 3D interaction focussing on projection based systems. We discuss the issues of collaborative 3D interaction using audio/video for face-to-face communication and the differences in using rear projection based Virtual Environments.

1. INTRODUCTION

The vision in Collaborative Virtual Environments (CVE) is to provide distributed teams with a virtual space where they can meet as if face-to-face, co-exist and collaborate while sharing and manipulating virtual data in real time. Therefore the environment needs to provide shared data representation, shared manipulation, integrate real-time video and audio communication and control between remote participants and at the same time provide a natural way of interacting with the shared data. For supporting the implementation and realization of such CVEs we report on our framework for the design and evaluation of distributed, collaborative 3D interaction focussing on projection based systems. The approach focuses on our CVE interaction taxonomy that supports the development of applications for small groups working together in rear projection-based VEs making use of video conferencing and 6DOF input devices. Design guidelines and the evaluation of different collaboration metaphors, operations, feedback components and user interfaces are also presented in the paper.

2. CVE INTERACTION FRAMEWORK

In order to find out how to support users we start with a very detailed *User's Task Description (UTD)*. A following *User's Task Analysis (UTA)* determines the so-called *User+Need Space (UNS)* which itself is the originator of the flow within our CVE taxonomy graph. The taxonomy can be found in earlier papers (Goebbels et al., 2000a; Goebbels et al., 2000b). This UNS relays the information extracted by the UTA of the UTD. We recommend to do an extensive, detailed description and analysis of the user's task in order to find out how the user's needs can be classified and addressed. Then the UNS deals with the following groups of issues:

- representation components, work mode, input/output device combinations, auxiliary tools as operations, metaphors and interaction techniques as well as actions and action feedback.

Representation Components are a very important part of Virtual Environments since they determine the representation of the visual parts of the application. The components are the representation of the user, the remote user, the environment, the virtual input device, the virtual tools and finally the representation of the data model and functionality.
The *Application+Interaction Space (AIS)* describes how users interact, with each other and collaboratively with the data set, in the virtual environment. In order to find the best interaction we first have to understand the low-level makeup of interaction. Therefore we have to split down interaction tasks and to find interaction templates which can be combined to form more complex interactions.
Awareness-Action-Feedback loops denote such interaction templates. These AAF loops allow us to understand and analyse very tiny steps in interactions. When analysing an interaction task of a single user with a data set we divide an autonomous AAF loop into four blocks. The first two blocks belong to the awareness phase where the user starts with proprioception (Mine et al., 1997). Proprioception allows the user to be aware of where s/he stands and looks

at, the position and orientation of body parts like arms, hands and fingers and everything that allows users to perceive themselves in relation to the environment. The next step is to be aware of the physical input devices held in the users hands and the virtual tool representations connected to them. The position and orientation of the virtual data set is perceived in this phase too. The user is then ready to perform an action. This action can for example simply be to move the hand together with the physical input device. After the action phase a feedback phase follows.
In this phase the user perceives the feedback from the action without which it is impossible to analyse the result of the action. In this case the user perceives the movement of the virtual tool representations as s/he moved the input device together with the hand. After the perception of the status of the situation the user can decide if the task is completed and therefore break the loop or whether the task is not completed yet and therefore prepare for the next action starting with the first block again.
Collaborative Awareness-Action-Feedback loops are of the same structure as the autonomous AAF loops. In addition to the autonomous AAF loops, the user perceives the co-presence during the awareness phase. It is comparable to proprioception but now information about the remote partner is queried. An interesting component represents the perception of co-knowledge and co-status. We found out that knowing that your partner is aware of you is one of the most important steps in this awareness phase. The user can confirm this status check either by voice or with the help of a gesture like the “thumbs up”. The action and the feedback phase are equal to the ones of the autonomous AAF loop. The Awareness-Action-Feedback loops are templates. With the help of operations, metaphors and interaction techniques described in (Goebbels et al., 2000b) it is now possible to give those templates a “face”. Depending on the user's subtask appropriate operations, metaphors and interaction techniques are chosen for each action.
We designed and implemented a medical CVE application according to the taxonomy and collaborative and autonomous AAF loops. We chose the most appropriate metaphors, operations, interaction techniques and representation components for this application. Two 2-sided Responsive Workbenches were used as the displays systems. The technical setup and an example from a real-time collaborative session are shown in Figure 1 (Goebbels et al., 2000a).

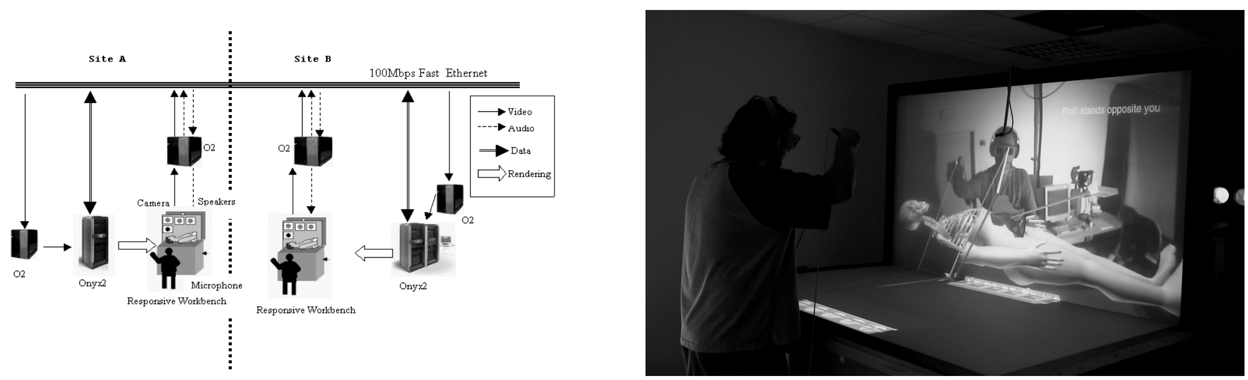

Figure 1. Built and used setup with two collaborative RWBs. Snap shot of a real-time collaborative session.

2.1 Evaluation

Three different evaluation methods are applicable when assessing Collaborative Virtual Environments (CVE). The expert heuristic, the formative and the summative evaluation (Hix and Hartson, 1993; Hix et al., 1999; Nielson, 1993). These evaluation methods make it possible to substantiate or refute realizations of a specific CVE. Assessing evaluators have to be no VE experts and not part of the developers team. Expert heuristic and formative evaluation are applied in alternating cycles in the early design state of the CVE. Based on the expert's knowledge, problems concerning usability can be solved following the expert's recommendations. After these recommendations are considered in a new and better design of the CVE the summative evaluation is applied. The objective of this evaluation method is to compare been different CVEs designed with the information obtained from the User+Need Space. Hence the output of the summative evaluation enables to statistically compare different realizations of interaction techniques, operations, representation components etc. and to choose the most appropriate one in terms of usability. However, important when planning an evaluation is to determine items which are assessable. This is often the most complex part. This collection of items is necessary to formulate specific questionnaires and hence to find and eliminate disturbance factors from the implementation of the CVE. For the assessment of the CVE the following factors are determined with respect to the User+Need Space defined by the User Task Analysis (see section 2.):

- menu representations
- virtual tool representations
- representation of data and its functionality
- environmental representations
- input devices
- physical equipment and cabling
- data processing and system reaction time
- graphical and acoustical resolution and quality
- network transfer rate
- perception of the own presence within the CVE
- perception of the partner's co-presence within the CVE
- perception of the collaboration in terms of equality of rights
- perception of the quality of collaboration
- frequency with which the user looked to the partner
- frequency with which the user spoke with the partner

Considering all these evaluation items in one session is almost impossible, since the items mentioned above evaluate too many different aspects of *Human-Computer-Human* interaction. In order to address this number of items special evaluation sessions are defined, namely the usability session, co-presence session and co-work session. An introduction is given prior to the evaluation sessions. During this introduction the evaluators are informed about the display system, the equipment and the environment they are going to work with. The objective is to create almost same conditions for all evaluators, since this is necessary for comparing numerical results of the formative and summative evaluations.

In the *usability session* the users (evaluators) interact autonomously within the VE for about five minutes.

During the interaction an external observer is taking notes and filling out a special observer questionnaire. This VE expert is observing the non-expert evaluator during the usability, the co-presence and the co-work sessions. Beside querying specific information about the time the user had to think and to debate before performing actions the questionnaire leaves space for informal observations. Especially this questionnaire helps to assess items which are difficult to be assessed by the evaluators themselves such like questions *"Did the user loose concentration during a session ?"* or *"How quickly could the user correct mistakes and continue the work ?"*. Information if the evaluator lost concentration during a session has an impact on the analysis and the way the numerical results have to be interpreted. However, this information can also imply the high cognitive load of interaction in the Collaborative Virtual Environment. Beside the overall ability to interact with the system critical incidents are very interesting to the observer.

In the *co-presence session* the user works again in the CVE but now with another data set. In contrast to the latter session an experienced user who has been involved in the development process is remotely present within the same environment through an audio/video connection. The experienced user explains the task, the data set, the input devices and the tools remotely to the evaluator. The remote partner who acts like a supervisor does not use any input devices or tools, but only gestures and verbal instructions. The task is to position three bones as precisely as possible to complement a human skeleton. These bones lay in front of the evaluator and look very similar to each other. If the evaluator does not know what to do the supervisor gives advise about the tools to be used, how to query information about the bones, how to change the viewpoint etc..

In the *co-work session* the task is slightly different. The task is to position three bones belonging to three different pairs to complement the human female skeleton collaboratively by both users. Each bone in a pair belongs to the left or the right side of the skeleton (i.e. the femur bone of the right and the left leg). A set of three of these bones lay in front of each user. As the users stand opposite each other, on different sides of the skeleton, they have to find out which bones belong to their side as the bones are mixed. Bones which belong to the partner's side can be exchanged by passing it over. To ensure further collaboration during the task the human female skeleton is covered by its skin. In order to position the bones, the particular part of the skeleton has to be made visible by cutting away the skin in this region. It is not possible to cut the skin permanently. This means that the cutting user holds the skin cutter while the other user positions the bone.

3. EVALUATION RESULTS

The expert heuristic, formative and summative evaluations for the different sessions delivered usability findings and recommendations.

The User+Need Space (UNS) for the considered evaluation scenario determines different representation forms for generic and content specific operations. For the generic operations a toolbar is designed whereas the content specific operations are grouped by a special ring menu (Goebbels et al., 2000b). In early designs of the CVE the generic toolbar was configurable in terms of its position by the user. The idea behind was that a dominant right-handed user might want to position the menu somewhere else in space than a dominant left-hander. Evaluation results showed that configuration of menus has a negative impact on the cognitive load. Additionally it is not really used in limited interaction spaces offered for example by the Responsive Workbench (RWB). Working with both hands at a RWB, the total viewing frustum is accessible in contrast to CAVE like display systems. Thus during the formative and summative evaluation the toolbar was positioned close to the users body within arm distance corresponding to the vendor's tray metaphor. Working at a RWB this toolbar is fixed whereas it is attached to the users body position when working in a CAVE or cylindrical and wall display systems.

Similar problems are encountered when using ring menus described in (Goebbels et al., 2000b). When a user intersects the data with the menu pick ray in the right hand the ring menu appears attached to the left hand and vice versa. This corresponds to the metaphor of handling a painter's palette with respect to dominant right and left-handers. The advantages were assumed to be the comfortable handling of this ring menu since it does not occlude any object being handled this way. For detaching the ring menu, over the shoulder deletion was integrated (Mine et al., 1997). Evaluation results showed that the handling making use of the painter's palette metaphor is not always as comfortable as assumed. The reason is that the user first has to recognize that the status of the hand changed as something is suddenly attached to it. Then the user has to look at the ring menu in order to select a content specific operation using the other hand. This is particularly annoying if the hand is busy with another task already. Additionally this metaphor makes it impossible to concentrate on the data set as the user is forced to turn the head towards the ring menu. In the improved design the ring menu is attached to the calling hand holding the menu pick ray. It follows the translation of the user's hand whereas the rotation of the user's wrist is used to intersect the ring pieces with the pick ray. The advantages are that the menu appears within the user's gaze and disappears as soon as the user releases the stylus button again. The menu is designed to be 70% transparent to avoid occlusion of data.

As already mentioned the menus group operations together. In order to apply operations tools are selected, e.g. the zoom operation requires a special zoom tool. The tools are represented by 3D icons which are attached to the buttons of the toolbar or to the choices of the ring menu. Usability findings showed that representations for the snap back tool, the information tool and the skin cutting tool were not appropriate in the early CVE design. Now the snap back tool is represented by a three dimensional hook icon, the information by a three dimensional “i” letter and the skin cutter tool by a three dimensional knife icon. These virtual tool representations increased the evaluator’s tool recognition rate by almost 80%.

Evaluation results indicated also that early approaches using two pinch gloves as input devices were not really addressing the user's needs. Reasons are the uncomfortable usage when working stand-alone collaboratively and trying to hand over pinch gloves to another user. Another encountered problem using pinch gloves together with pick rays is that it is almost impossible to keep pointing somewhere and additionally snap with the middle finger and the thumb for selection. Similar problems using pinch gloves have been encountered in (Hix et al., 1999). Improvements are made by using a special three button tool in one hand and a stylus in the other. The reason for not using three button tools in both hands refers to the high cognitive load of their usage due to the many buttons. After modification evaluation showed that the stylus is rather used in the dominant and the three button tool in the non-dominant hand.

A *sharing viewpoint* metaphor is implemented for manipulating the users’ viewpoint (Goebbels et al., 2000b). Evaluation results showed that an exo-centric viewpoint manipulation is better than an ego-centric when *standing almost beside the partner*. In this context exo-centric manipulation is based on how a user would act in real world by moving laterally. When sharing the same viewpoint *(looking through the partner's eyes)* or sharing the mirrored viewpoint *(looking from opposite the partner)* ego-centric viewpoint manipulation is implemented. This manipulation is realized by pressing and releasing a special button on the three button tool. These observations are valid working at a Responsive Workbench. Because of the limited interaction space it is possible to access the data set visually from all sides by manipulating the viewpoint as described above. However, other own evaluations showed that in the CVE implemented using a CAVE and a cylindrical display no ego-centric viewpoint manipulation is needed. Here users prefer exo-centric viewpoint manipulation due to the larger interaction space and the perception of entire immersion.

In the co-work session the evaluators complement a female skeleton by missing bones. There the task is aggravated as the skin of the body is cut in order to make the skeleton visible. Usability findings indicated that users prefer to get a quick overview of the situation. This leads to the implementation of a content specific wireframe operation. The users are able to only render the skin of the body in wireframe and thus have a direct view onto the underlying skeleton. With this strategies can be discussed and collaborative tasks can be planned more quickly. This content specific wireframe operation is only usable for getting an overview. For complementing the skeleton the skin has still to be cut.
In addition to that observations of critical incidents during the co-presence session are made. These critical incidents occur due to network drop outs, indicating that the perception of co-presence is interrelated with the video frame rate. Further experiments with the video frame rate as parameter showed that the perception of co-presence vanishes completely if the video frame rate sinks below 12 fps.

4. CONCLUSIONS

We presented our interaction taxonomy for designing and creating Collaborative Virtual Environments. They provide distributed collaborative teams with a virtual space where they could meet as if face-to-face, coexist and collaborate while sharing and manipulating virtual data. Further we discussed the issues involved in bringing together Human Computer Interaction and Human to Human Communication, focusing on projection-based Virtual Environment systems. Evaluation result derived from alternating cycles of expert heuristic and formative and summative evaluations are also discussed.

The work reported was supported by the Humboldt-University of Berlin and the German Ministry of Research and Technology (BMBF) under grant number 01KX9712/1.

REFERENCES

Goebbels, G., Aquino, P., Lalioti, V., and Goebel, M. (2000a). Supporting team work in collaborative virtual environments. In Proceedings of ICAT 2000 - The Tenth International Conference on Artificial Reality and Tele-existence.

Goebbels, G., Lalioti, V., and Goebel, M. (2000b). On collaboration in distributed virtual environments. In The Journal of Three Dimensional Images, Japan, 14(4) , pages 42-47.

Hix, D. and Hartson, H. R. (1993). User interface development: Ensuring usability through product and process. New York: John Wiley and Sons.

Hix, D., II, E. S., Gabbard, J. L., McGee, M., Durbin, J., and King, T. (1999). User-centered design and evaluation of a real-time battlefield visualization virtual environment. IEEE, pages 96-103.

Mine, M., Frederick, P., Brooks, J., and Sequin, C. (1997). Moving objects in space: Exploiting proprioception in virtual environment interaction. Proceedings of SIGGRAPH 97, Los Angeles, CA.

Nielson, J. (1993). Usability engineering. Academic Press.