Collaboration in Co-located Automotive Applications

by Holger Salzmann

A dissertation submitted in conformity with
the requirements for the degree of Dr.-Ing.

Bauhaus-Universität Weimar
Fakultät Medien

Abstract

Virtual reality systems offer substantial potential in supporting decision processes based purely on computer-based representations and simulations. The automotive industry is a prime application domain for such technology, since almost all product parts are available as three-dimensional models. The consideration of ergonomic aspects during assembly tasks, the evaluation of human-machine interfaces in the car interior, design decision meetings as well as customer presentations serve as but a few examples, wherein the benefit of virtual reality technology is obvious. All these tasks require the involvement of a group of people with different expertises. However, current stereoscopic display systems only provide correct 3D-images for a single user, while other users see a more or less distorted virtual model. This is a major reason why these systems still face limited acceptance in the automotive industry. They need to be operated by experts, who have an advanced understanding of the particular interaction techniques and are aware of the limitations and shortcomings of virtual reality technology.

The central idea of this thesis is to investigate the utility of stereoscopic multi-user systems for various stages of the car development process. Such systems provide multiple users with individual and perspectively correct stereoscopic images, which are key features and serve as the premise for the appropriate support of collaborative group processes. The focus of the research is on questions related to various aspects of collaboration in multi-viewer systems such as verbal communication, deictic reference, embodiments and collaborative interaction techniques.

The results of this endeavor provide scientific evidence that multi-viewer systems improve the usability of VR-applications for various automotive scenarios, wherein co-located group discussions are necessary. The thesis identifies and discusses the requirements for these scenarios as well as the limitations of applying multi-viewer technology in this context. A particularly important gesture in real-world group discussions is referencing an object by pointing with the hand and the accuracy which can be expected in VR is made evident. A novel two-user seating buck is introduced for the evaluation of ergonomics in a car interior and the requirements on avatar representations for users sitting in a car are identified. Collaborative assembly tasks require high precision. The novel concept of a two-user prop significantly increases the quality of such a simulation in a virtual environment and allows ergonomists to study the strain on workers during an assembly sequence. These findings contribute toward an increased acceptance of VR-technology for collaborative development meetings in the automotive industry and other domains.

Zusammenfassung

Virtual-Reality-Systeme sind ein innovatives Instrument, um mit Hilfe computerbasierter Simulationen Entscheidungsprozesse zu unterstützen. Insbesondere in der Automobilbranche spielt diese Technologie eine wichtige Rolle, da heutzutage nahezu alle Fahrzeugteile in 3D konstruiert werden. Im Entwicklungsbereich der Automobilindustrie werden Visualisierungssysteme darüber hinaus bei der Untersuchung ergonomischer Aspekte von Montagevorgängen, bei der Bewertung der Mensch-Maschine-Schnittstelle im Fahrzeuginterieur, bei Designentscheidungen sowie bei Kundenpräsentationen eingesetzt. Diese Entscheidungsrunden bedürfen der Einbindung mehrerer Experten verschiedener Fachbereiche. Derzeit verfügbare stereoskopische Visualisierungssysteme ermöglichen aber nur einem Nutzer eine korrekte Stereosicht, während sich für die anderen Teilnehmer das 3D-Modell verzerrt darstellt. Dieser Nachteil ist ein wesentlicher Grund dafür, dass derartige Systeme bisher nur begrenzt im Automobilbereich anwendbar sind.

Der Fokus dieser Dissertation liegt auf der Untersuchung der Anwendbarkeit stereoskopischer Mehrbenutzer-Systeme in verschiedenen Stadien des automobilen Entwicklungsprozesses. Derartige Systeme ermöglichen mehreren Nutzern gleichzeitig eine korrekte Stereosicht, was eine wesentliche Voraussetzung für die Zusammenarbeit in einer Gruppe darstellt. Die zentralen Forschungsfragen beziehen sich dabei auf die Anforderungen von kooperativen Entscheidungsprozessen sowie den daraus resultierenden Aspekten der Interaktion wie verbale Kommunikation, Gesten sowie virtuelle Menschmodelle und Interaktionstechniken zwischen den Nutzern.

Die Arbeit belegt, dass stereoskopische Mehrbenutzersysteme die Anwendbarkeit virtueller Techniken im Automobilbereich entscheidend verbessern, da sie eine natürliche Kommunikation zwischen den Nutzern fördern. So ist die Unterstützung natürlicher Gesten beispielsweise ein wichtiger Faktor und es wird dargelegt, welche Genauigkeit beim Zeigen mit der realen Hand auf virtuelle Objekte erwartet werden kann. Darüber hinaus werden Anforderungen an virtuelle Menschmodelle anhand einer Zweibenutzer-Sitzkiste identifiziert und untersucht. Diese Form der Simulation, bei der die Nutzer nebeneinander in einem Fahrzeugmodell sitzen, dient vor allem der Bewertung von Mensch-Maschine-Schnittstellen im Fahrzeuginterieur. Des Weiteren wird das neue Konzept eines Mehrbenutzer-Werkzeugs in die Arbeit mit einbezogen, da hier verdeutlicht wird wie die Simulation von Montagevorgängen in virtuellen Umgebungen mit passivem haptischem Feedback zu ergonomischen Verbesserungen entsprechender Arbeitsvorgänge in der Realität beitragen kann. Diese Konzepte veranschaulichen wie VR-Systeme zur Unterstützung kollaborativer Prozesse in der Automobilbranche und darüber hinaus eingesetzt werden können.

Acknowledgments

Many people were involved in the process of helping me make this thesis become a reality. I want to take this opportunity to say thank you to all who helped and mention some of them here. Bernd Fröhlich introduced me to the idea of multi-viewer stereo displays and provided guidance, advice and motivation. Martin Bergmann opened up the opportunity for me to research this thesis at Volkswagen AG and supported me in gathering the necessary resources, while also offering helpful advice. I want to thank my students, Lourdes Sanchez Paricio and Manuel Gleisberg. Without their help only a fraction of this work could have been investigated and realized. The VR Systems Group within the Media Faculty also need to be thanked for their valuable discussions and support in setting up advanced hardware devices necessary to operate the multi-viewer system. I would also like to extend my gratitude to the Virtual Reality Lab of Volkswagen Group Research, especially Ralf Rabätje, Mathias Möhring, Ralph Voges and Jan Jacobs for sharing resources and supporting me with the studies. Finally, I thank my parents and Linda for their love, patience and support.

Contents

Chapter 1

Introduction

In the development processes of the automotive industry a growing number of virtual car-models is used instead of real prototypes. While CAD-software and other simulation tools are already well established, virtual reality technology still faces only limited acceptance among engineers and managers, and thus it is only used for a few niche applications. This is the case even though there exists a number of predestinated application areas for such visual 3D simulations. The evaluation of ergonomic aspects and the human machine interface of the car interior, in product creation as well as for customer presentations serve as examples. In almost all these areas, a group of people is involved in the task (e.g. typical development meetings) wherein virtual simulations serve as a communication support medium. Such meetings are usually attended by a group of experts from different automotive fields and departments (Figure 1.1).

However, the automotive industry has only employed single-user VR-technology so far. Head-mounted display (HMD) systems isolate the user from the surrounding environment and his colleagues, which makes collaborative discussion impossible. The regular projection-based stereoscopic systems only calculate the image for a single tracked user or for an average viewpoint, which results in a distorted perception of the 3D model if the user is not in the location the image was calculated for. Despite high demands on visual quality and haptic feedback, most of all, the insufficient support of multiple users remains the reason for a continued low acceptance of virtual reality technology in automotive group meetings.

The central idea of this thesis is to investigate the utility of stereoscopic multi-user systems for various stages of the car development process. Such systems provide multiple users with individual and perspectively correct stereoscopic images, which is a key feature for the appropriate support of collaborative group processes in 3D environments. However, the introduction of stereoscopic multi-viewer systems to the automotive context also poses a number of challenging research questions:

- What are the appropriate multi-viewer tools and interaction techniques to enable true collaboration among automotive experts?
- Which precision is required to enable natural gesturing and pointing? How is it possible to achieve the required precision?
- Which scenarios are appropriate for which kind of multi-viewer technology?
- Are avatar representations for the users necessary? Which avatar fidelity is required?
- How does verbal communication influence collaboration in co-located multi-user setups?

To answer these and other questions, a number of different automotive scenarios were identified, in which the support for multiple users is most pressing. Design reviews and assembly simulations require full scale representations of the car and always involve a group of experts. Current seating buck implementations enable only one person sitting in a car mockup to evaluate ergonomics of new human machine interfaces of the car interior. However, in real car prototypes, two or more experts enter the car for discussion. These scenarios have different requirements with respect to the appropriate display technology, number of involved users, support for interaction and other factors, which are presented in the following sections.

1.1 Seating Bucks

The automotive industry uses physical seating bucks, which are minimal mockups of a car interior, to assess various aspects of the planned interior early in the development process. In a virtual seating buck (Figure 1.2), users wear a head-mounted display which overlays a virtual car interior on a physical seating buck. This system is used for the development, testing and evaluation of novel human-machine interface concepts for future car models.

Virtual seating bucks allow the evaluation of visibility and reachability aspects in the car interior from the driver's viewpoint. A seating buck is basically comprised

Figure 1.1: Typical project discussion in front of a real prototype.

of a driver seat, a steering wheel and the pedals that are jointed together on a platform. Such physical seating bucks are repeatedly used in different steps of the car development process to evaluate ergonomics of the car interior. Unfortunately, it is common practice to build a new seating buck for each kind of problem, which in turn has a direct impact on development costs. In contrast, a virtual seating buck offers the opportunity to visualize different concepts in only one setup, thereby reducing development efforts. A basic setup of real car parts is used to mediate the feeling of sitting in a real car to the user by providing passive haptic feedback. The virtual car and its surroundings are completely generated by the HMD. The limitation is that only the user wearing the HMD is able to perceive a stereoscopic virtual car and no other users standing nearby.

For some time, single-user seating bucks were used to explore the qualities and requirements of new human-machine interfaces in cars. An initial survey was performed with 28 participants of the three potential user groups for the virtual seating buck technology: electronics developers, interface designers and ergonomists. The goal was to get a prioritization of the next steps which are required to increase the acceptance of the seating buck in the development process. All participants of

Figure 1.2: The single-user seating buck.

the aforementioned user groups had to complete a certain set of tasks while seated in the seating buck and wearing the HMD. They were asked to perform tasks such as:

- individually adjusting the steering wheel and seat in a comfortable way with the help of the remaining mechanics physically available
- taking a look at the backseat by turning around
- looking through the windshield and looking for the upper lamp of the traffic light
- turning the steering wheel
- switching on the cd-playback of the radio
- entering a route into the navigation system

The participants were required to answer questions while seated, as well as in a post-questionnaire immediately following the virtual session. The questions addressed analyzed realism, interaction performance, comfort, attitude toward the technology and general remarks.

The results of the study revealed that the limited haptics and the imperfect rendering quality are major obstacles toward acceptance, but it was also mentioned that the immersed user wearing the HMD is partitioned from people surrounding him. Some of the participants directly asked if it is possible to integrate a second user as the co-driver. Due to the single-user nature of the regular seating buck, other users had to watch the actions of the driver on a regular monoscopic projection screen, which was set up at quite a distance from the seating buck. Thus communication between the passive observers at the large screen and the driver in the seating buck proved especially difficult.

The idea was to build a two-user seating buck to introduce a second expert into the environment of the virtual seating buck to enable face-to-face discussions of car interface concepts as in the corresponding real-world scenario. The development, implementation and evaluation of a two-user seating buck is described and the requirements for this co-located interaction scenario are identified.

1.2 Assembly Simulations

For many years now, immersive virtual environments have been employed in certain fields of the automotive industry. In the area of assembly simulations, they are used for two different aspects: testing and optimization of the assembly processes; and the assessment of the ergonomics of an assembly sequence. In both situations, the display of a full scale model and the involvement of multiple people are crucial for the validity and significance of these evaluations.

In the context of automotive assembly simulations, typically a group of engineers meets in front of a large projection screen. They discuss how certain parts can be assembled on the car body. Therefore, they have to clarify if collisions with other parts can occur or the accessibility by special tools is given. In most cases, the car model is displayed in original size on the projection screen. Usually a VR-expert assists the group, keeps the goal of the meeting in focus and is able to answer pending questions regarding VR-related issues. He is also the only one who

Figure 1.3: One tracked user and one non-tracked user in front of an L-shaped projection screen.

has a correct stereoscopic view and is in control of the input device as well as all implemented virtual functions.

Regarding projection-based stereoscopic systems used in automotive applications, it is essential that more than one user can join the virtual presentation on the projection screen by putting on his stereo-glasses. But in commonly used setups only one user, the master, has a correct stereoscopic view since his glasses are tracked by a tracking system. Due to the tracking, his stereoscopic view is always processed correctly depending on the position and viewing direction of his head and glasses. The other users are equipped with stereo-glasses as well, but since they are not tracked their view would only be stereoscopically correct if they had the same position as the master (Figure 1.3). The more they are standing apart from the master, the more their view on the virtual scene is distorted (Figure 1.4).

Additionally, the scene becomes distorted depending on the master's head movement. Even simple interactions like pointing at certain car parts become impossible considering that already two users would point at different positions in the virtual space due to the distortions. So again, as previously mentioned in the seating buck

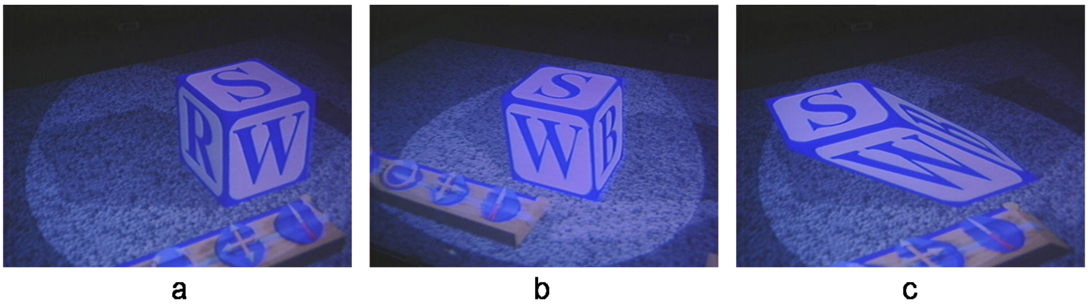

Figure 1.4: (a) The view seen by a user standing on the left side in front of the projection screen. (b) The view seen from the right side. (c) The view a user would see standing on the left side in front of the projection screen if generated from the position of the view on the right. Even though the image on the screen is the same in (b) and (c), the cube appears sheared to the non-tracked user in (c). (by Agrawala et al. [1997])

paragraph, automotive users are requesting the support of two or even more users to participate in a virtual scenario. Within this thesis interaction techniques for co-located collaboration in projection-based two-user systems are developed and evaluated.

1.3 Side-by-Side and Face-to-Face Collaboration

Common collaborative scenarios involve users standing side-by-side or face-to-face and require direct interaction within the reach of the users' arms. For side-by-side interactions of two workers, such as standing in front of a car and mounting engine parts, the workers employ different tools or they assume different roles. Automotive experts are also interested in looking at other tasks; more specifically, tasks wherein users have to work face-to-face. These tasks can only be supported by an HMD-based two-user system since any projection-based setup would visually result in problematic occlusions if an object would have to be displayed between the two users. True face-to-face interactions (e.g. wherein two workers are collaboratively manipulating an object) are developed and evaluated to assess the suitability of different co-located immersive virtual reality solutions and appropriate interaction forms for these types of tasks.

1.4 Outline

In current VR-setups used at the Volkswagen laboratories, there is only weak support for the basic activities of collaborative work in a co-located group of users. The focus is often on advanced, abstract and indirect interaction with virtual models which might indeed be helpful, but are only understood by experienced users of VR-systems. Since all the systems are designed for a single user only, it is difficult to show novice users how to interact with a VR-system. Multi-user systems allow experts to show novice users how to use the system, and they enable more direct and real world-like interaction based on gesturing and pointing. In the scope of this thesis, five aspects play a major role when designing and evaluating collaborative co-located virtual environments in the context of automotive scenarios:

- vision - individual stereoscopic views
- speech - communication with others by natural voice
- referencing - by pointing gestures
- embodiment - perception of users' bodies through suitable avatar representations
- collaborative interaction - supported by passive haptic feedback and real world props

The remainder of this thesis is organized as follows. In Chapter 2 similar work is reviewed and related to this thesis.

Chapter 3 describes how a two-user projection-based setup is implemented enabling side-by-side collaboration. An initial study comparing collaboration in a common single-user setup to a new two-user setup is performed. The intention of this study is to investigate the advantage of a stereoscopic two-user system over a conventional single-user solution for common automotive scenarios and how it extends collaborative actions. The usability of the two-user display is investigated in an expert review involving automotive experts.

In Chapter 4 a basic interaction metaphor - pointing with the real hand - is investigated in a two-user projection system. In real life, pointing is an indispensable part of communication and occurs automatically during conversation. The question is how accurate is pointing in virtual environments in comparison to real-world pointing and if the type of pointing does indeed affect accuracy. In an experiment, the accuracy of the users' ability to point at virtual objects with their real hand is investigated. How well users can identify objects that are pointed at with a real hand by others is also examined.

Chapter 5 is focused on face-to-face collaboration in HMD-based automotive scenarios. In a virtual assembly task, two users are enabled to simultaneously manipulate one object. Two interaction techniques are compared with respect to their usability in a collaborative task. Since the task is to observe ergonomics of a certain assembly process, it is investigated which technique provides sufficient accuracy and enhances interaction and coordination between users.

Chapter 6 introduces the two-user virtual seating buck which is a great application area for investigations on virtual embodiments. Considering that this application is focused on the development of new human machine interfaces of the car interior, its main purpose is the effective and intuitive manipulation of such interface elements. Thus a basic user avatar might be sufficient for the evaluations of new interface concepts. The question is how basic such a representation can be in order to be acceptable for automotive users such that the manipulations are perceived as being realistic.

Finally, Chapter 7 concludes this thesis by summarizing the contributions, discussing conclusions and presenting directions for future research.

Chapter 2

Related Work

This thesis focuses on co-located collaborative virtual environments (CVEs) wherein people are provided with visual information displayed in either projection-based or HMD-based systems. Direct verbal communication is possible since users are co-located and they can reference objects surrounding them with their hands and arms. The awareness of collaboration and coordination of actions between users is enabled by appropriate body representations. To support collaborative processes various interaction techniques are introduced and evaluated.

2.1 Computer Supported Cooperative Work

Computer Supported Cooperative Work is an interdisciplinary field of research focusing on the collaboration of groups with the help of computers. The goal is to increase the efficiency of collaboration by providing information via suitable communication technologies.

According to Rodden [1992] cooperative systems can be classified by looking at two aspects; the location of the users and the mode of cooperation. The mode of cooperation can be either synchronous or asynchronous. While synchronous cooperation requires the simultaneous presence of all collaborating users, like in meeting rooms, asynchronous applications include messaging and co-authoring systems. As a general definition, a meeting room consists of a large screen and several computer terminals. Such configurations support face-to-face collaboration between users in a local group and are often classified as decision conferences. In a more recent work Rama and Bishop [2006] define CSCW as the study of how people use technology, with relation to hardware and software, to work together in shared time and space. They propose a classification of time as being either synchronous or asynchronous, and space as being either distributed or co-located. All applications introduced in the thesis at hand are synchronous in time and support the collaboration of co-located users.

Tan et al. [2001] suggest the use of Mixed Reality (MR) to best support face-to-face collaboration. MR is the combination of physical real world objects and computer generated virtual information. In real world verbal communication, gestures, body posture and facial expressions provide social awareness of other users and therefore enrich collaborative interaction. People have individual views on the real space but manipulating an object, handing it over or pointing and looking at it are basic interactions they all share. Tangible, real world objects allow simultaneous manipulation by more than one user, and they are easily reconfigured. Common virtual environments remove users from the real world and consequently, they lose important social cues. With the basic interactions mentioned above, people should be able to use objects as spatial references as in the real world. Tan et al. [2001] determine that the choice of display technology directly influences the interaction methods and interface design. They differentiate between transparent HMDs and a large projection screen. In both configurations users can still perceive the surrounding as the real world. The drawback of the HMDs is that they are obtrusive devices but support a fully surrounding virtual environment. The projection screen provides higher levels of comfort but limits the workspace to only a small frame. Despite the choice of display technology, the manipulation of physical objects to control corresponding virtual objects is an intuitive interaction metaphor for individual and shared workspaces. At the same time, such interactions greatly provide visual, tactile and social feedback concerning other users.

Mandryk et al. [2002] identified seven physical display factors such as orientation, size and number of displays that influence co-located collaboration in groups. In detail, they suggest a categorization to be considered in the following:

- **Orientation of Display** - Displays can be oriented vertically like projection screens or horizontally like tabletop displays. With vertical displays all users nearly have the same view on the scene while with horizontal displays, users can have opposite views.
- **Arrangement of Users at the Display** - Whether the users are positioned side-by-side or face-to-face determines the impact on the interaction with the

display and with each other. Users facing each other can see the display and others at the same time, which enables direct eye contact and a variety of gestures. If the users are positioned side-by-side these possibilities are clearly limited but other features of collaboration may arise.

- **Size of Display** - The larger a display is the more users can participate in the visualized content. This facilitates work-sharing since users can work on different subtasks simultaneously.
- **Proximity to the Display** - In larger groups it is the case that some users are relatively far away from the display and therefore treat it as non-interactive. If the display and interaction space is in arm's reach there is a higher motivation for users to interact with the content by directly referencing objects.
- **Privacy of the Display** - Large displays that allow shared viewing within a group of people enable better awareness of activities and information. Privacy can be increased if all users share the same scene but are provided with additional specific information regarding their expertise.
- **Superimposition of Display Space on the Input Space** - If display and input space are congruent, there is an increased awareness of actions between collaborating users. Arm movements, for example, can be perceived in peripheral vision in contrast to movements of a small cursor.
- **Number of Displays** - One large display, such as a projection screen, enhances collaboration in a group of users. If each user is provided with an individual display, such as handhelds or laptops, it may be that they are focusing only on a small part of the workspace. If multiple displays mean multiple HMDs, this might not necessarily be the case.

These factors of physical display technology influence co-located collaboration and provide an understanding of how various technical features can form a basis to compare different collaborative systems. Furthermore, a conclusion is that the more users are immersed in a CVE, the further enhanced the performance of collaboration is. The thesis at hand focuses on co-located collaboration by using both projection-based and HMD-based setups. Regarding the *orientation of displays* the projection screen is vertically oriented but the HMD provides a full surround view and is not restricted vertically or horizontally. The *arrangement of users at the display* plays a major role in the context of this thesis since collaborative interactions are distinguished in both side-by-side (projection screen) and face-to-face (HMD) scenarios. In both conditions, display space and input space are superimposed. With respect to the factors *number of displays*, *proximity to the display* and *size of display* HMDs are an exception since the displays themselves are very close to the eye and therefore offer a wider view of the whole workspace.

2.2 Collaborative Virtual Environments

Collaborative virtual environments (CVEs) support the interaction of multiple remote or co-located users in a shared virtual environment. The hardware setups introduced in this thesis share similarities with other co-located setups like the Studierstube project by Schmalstieg et al. [2002], the PIT by Arthur et al. [1998] and the Two-User Responsive Workbench by Agrawala et al. [1997]. Two or more users are co-located in the same room and experience the same virtual scene. Each of these systems provides an individual stereoscopic view for each user, supports head tracking and the perception of the surrounding real environment.

Figure 2.1: The PIT by Arthur et al. [1998].

The PIT enables two users to concurrently experience the same virtual scene in a dual-projection-screen stereo display while sitting next to each other (Figure 2.1). In addition to visual information provided by the virtual environment, the system supports verbal communication and gestures. The main goals in designing the PIT were to create a high quality 3D display, to give access to common devices, to support work within arm's reach and to include a second user. In their previous work Arthur et al. [1998] observed that their scientific users commonly work together in pairs. According to this, they wanted to provide a second user with an individual stereoscopic view equal to that of the first user. Other solutions such as using only a monitor for the second user were not satisfying and did not solve the problem of a distorted view due to the lack of self-controlled head tracking.

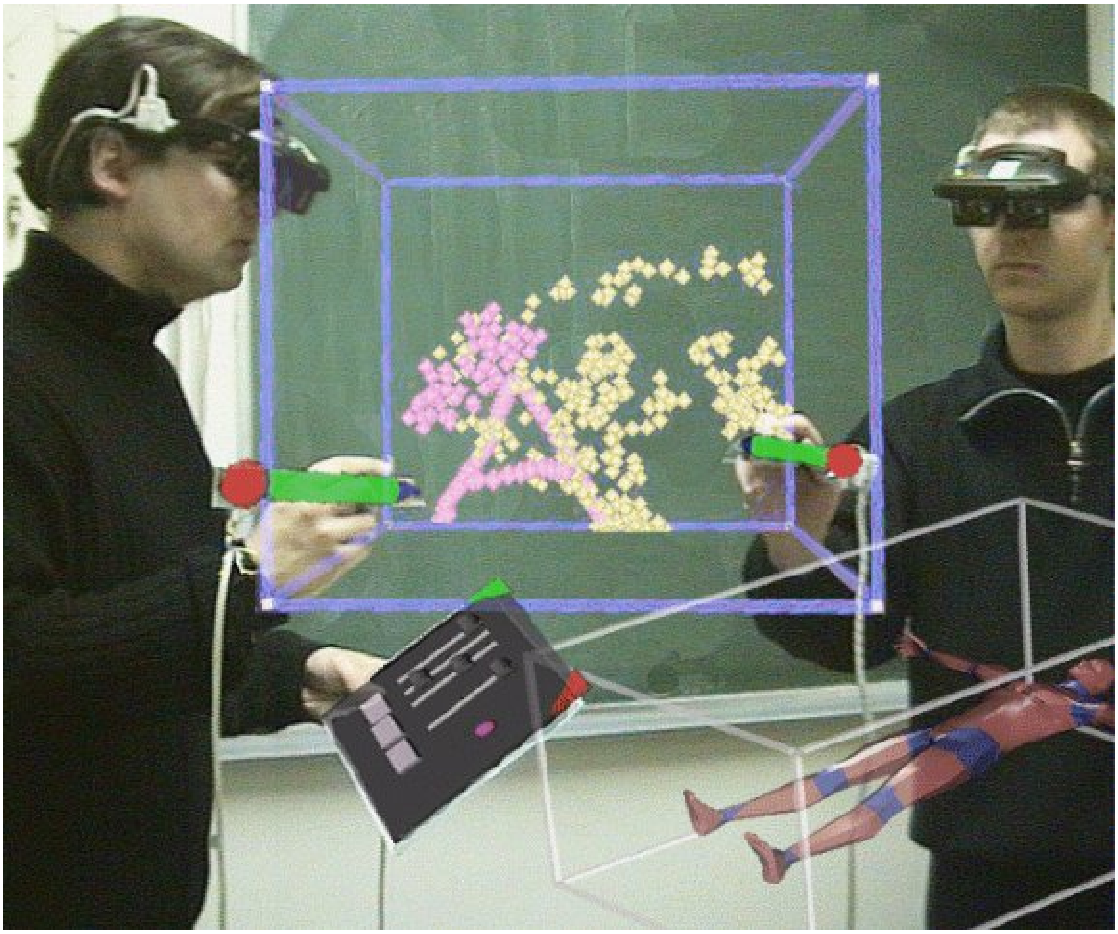

Figure 2.2: The Studierstube augmented reality project by Schmalstieg et al. [2002].

The Studierstube system implements collaborative augmented reality by augmenting the real work environment with virtual images. Two handheld devices, a panel and a pen, are used to interact with the system. This so-called personal interaction panel greatly facilitates intuitive handling since it has a familiar shape and supports passive haptic feedback. It is restricted to be used by only one user and is not supposed to support simultaneous collaborative interaction between two users. The main purpose of the panel is to control the application out of the virtual environment without abandoning immersion. Therefore, it is augmented with task-specific information (e.g. sliders to manipulate a cutting plane). Each user is provided with an individual stereoscopic view by wearing a tracked, see-through HMD (Figure 2.2). The collaborative workspace is not limited to two users. See-through HMDs allow the users to perceive their real environment and enable interactions with objects as well as face-to-face communication with co-located users through verbal communication and gestures.

Figure 2.3: The two-user responsive workbench by Agrawala et al. [1997].

In contrast to other setups, the two-user tabletop responsive workbench introduced by Agrawala et al. [1997] is a projection-based display system. Both users are looking at the same screen that generates two individual stereoscopic views by using a frame-interleaving technique. The workbench approach allows two users to interact in a side-by-side manner by standing together on one side of the bench (Figure 2.3), but it also enables face-to-face collaboration between the users standing adverse to each other with the workbench in between. In this way, the users are able to see their counterpart through the stereo glasses and experience direct verbal communication supplemented by lip motions and gestures. The drawback is that when looking into each other's faces, the virtual environment on the workbench cannot be perceived anymore, but hand gestures such as pointing at virtual objects on the bench can be perceived very well without abandoning the view at the virtual scene.

Referencing to objects through the use of pointing gestures is an important part of communication both in reality and in collaborative virtual environments. Wong and Gutwin [2010] state that it is more complex to point in CVEs. They investigate how well users can identify pointing targets in CVEs by using different pointing techniques because not all types of pointing at certain objects require high accuracy. In an initial study, they identify several ways of pointing and in a second study they want to find out how accurately users can produce and interpret pointing directions. Deictic pointing offers the opportunity to identify an object by simple gestures instead of complex verbal descriptions. Intuitively, people are experts in producing and interpreting pointing gestures in the real world and pointing is closely coupled with verbal utterances. One of the most difficult things in CVEs is to imagine another user's view, and the limited field of view of most CVEs reduces the ability of observers to perceive pointing gestures or viewing directions of others. Another problem is that pointing is often triggered by symbolic or spoken commands and not by the arm of the user or by an avatar representation. Furthermore, visual and perceptual factors like limited depth cues, low resolution, artificial avatars and reduced field of view make real and virtual pointing more different. In co-located setups, pointing gestures are important to direct a group's attention to a certain spot (e.g. automotive development meetings). In their experiments Wong and Gutwin [2010] compare three different pointing techniques: pointing gesture only, pointing plus speech and pointing plus written notes. Altogether, the gestures were more complex without additional information on the communication channel. In the real world condition, users pointed at a monoscopic projection screen with a switched off laser pointer that was switched on later to determine if the target was hit. In the virtual condition, users let a virtual avatar point at a target controlled by a 2D-mouse. In both conditions, the observing users tried to identify the target from two positions, behind and beside the pointing user.

Wong and Gutwin [2010] conclude with a list of new design requirements regarding deictic pointing as well as confirming already known items:

- Deictic pointing has varying accuracy requirements.
- CVEs should support multiple types of pointing.
- Reduced communication richness in CVEs may increase requirements for accurate pointing.
- The importance of peripheral vision.
- Natural pointing in CVEs can be successful.
- Pointing in CVEs is still less accurate than in the real world.
- Compressed field-of-view does not aid accuracy.

The pointing gestures in their study were controlled by commands or mouse-events. It will be interesting to find out how more natural gestures would affect referencing in CVEs, such as pointing with a real hand and arm or with a virtual representation mapped to the real body part's movement (cf. Chapter 4).

Duchowski et al. [2004] suggest another method to implement deictic referencing in CVEs. They investigate the influence of visual deictic reference in a scenario with two users immersed in the same virtual environment. By tracking head or eye movement, they generate a colored virtual lightspot that is either head-slaved or eye-slaved. It is compared whether a head-slaved or eye-slaved reference is more helpful to disambiguate certain points of reference. The co-located setup includes two HMDs, one of which is equipped with an eye-tracking device. The users are standing six feet apart in a virtual room with road signs printed on the walls as reference targets. In several trials the referee had to verbally identify the road sign the referrer was aiming at. In summary Duchowski et al. [2004] presented a method to display visual deictic reference in a CVE. The lightspot representation can be compared to the dot of a laser pointer emerging from the user's eye or head. The results show that an eye-slaved lightspot is more effective than a head-slaved lightspot, but both methods are beneficial when identifying referenced objects. The authors assume that instead of the lightspot, it may be equally effective to implement expressive avatars supporting head and eye movement. They further recommend to provide articulated avatars as possible options, mapping movement of head, eyes, arms and torso according to tracking capabilities. Visual deictic reference is another method to refer to objects in a CVE, but well articulated avatars with rotating heads and eyes can only be implemented in a setup with non-see-through HMDs or in a distributed setup. Otherwise the virtual eyes will be overlaid with the real body of the users.

Hindmarsh et al. [2000] describe that it is hard for users in a CVE to imagine their counterpart's perspective. It is often the case that users are not able to determine what another user refers to, either by looking or by pointing. In CVEs, a virtual embodiment provides users with information of others and their current focus of interest. As in the real world, embodiments can look and point at virtual objects surrounding the users. Therefore, it is necessary that the embodiment is at least humanoid with head, torso and limbs. The embodiment implemented by Hindmarsh et al. [2000] supported pointing with the arm as well as inclining the head towards the target. For manipulation tasks, the humanoid embodiment was extended by drawing a connecting line between arm and the object that should be moved. The purpose was to visualize that objects could be manipulated over a distance. In their experiment, it occurred that manipulated objects and the manipulating body part were not displayed in the same image due to a limited field of view (55 degree). This was intuitively compensated by users as they verbally described their current actions in more detail to the other user. Even with the embodiment in their view, the users tend to falsely predict the other's perspective, orientation and focus of interest. Due to the limited field of view, the workspace is fragmented. For example, if one user references an object by speech, "this object there", and points at it, the other user looks in his face, then at his pointing arm and finally at the object. In that way, face, arm and object cannot be concurrently in the same image. Hindmarsh et al. [2000] state that it is problematic for users to reassemble relations between object and body, or in other words, to retrieve the referent. It was furthermore observed that users turn away from an object to find the referent's gesture only to immediately return to face the object again. Or the users have to extensively search the pointing arm to find the relevant object. In reality, these steps are only short glances, but in a CVE with limited field of view, this process requires extensive and repeated head movements to allow the users to orient towards the object. So the preparation of a manipulation task can take longer than the manipulation itself. This has a strong impact on the dynamics of collaboration in a group since users spend a lot effort in describing objects and preparing each other instead of focusing on the next step (e.g. where to place an object). The workflow of a collaborative process and accompanying discussions are therefore disrupted. In co-located collaborative work, the coordination of actions is enabled by the ability to monitor the activities of others. A limited field of view constrains the peripheral awareness of those activities. For example, if one user sees another user in a CVE, he assumes that the other is able to see his gestures as well, but in fact he cannot see them. It is hard to imagine another's perspective in the CVE. In reality, users rely on their visual perception of others, but if this information is constrained they need to reveal each other's actions by increased verbal communication. To compensate such limitations Hindmarsh et al. [2000] recommend four items that should be addressed in CVEs:

- Limited field of view makes it difficult to simultaneously perceive source and target of actions (e.g. pointing). Confusion among users arises by assuming incorrect field of view of others.
- Limited information about other's actions both on embodiment and target objects.
- Limited movement capabilities in a CVE due to problems of locating objects and reduced system performance.
- Limited parallelism for actions since it is difficult to interact concurrently (e.g. looking and pointing or grasping and moving).

The authors also suggest general solutions to overcome these limitations. One recommendation is to use more immersive displays like HMDs or a CAVE™instead of desktop displays. HMDs are supposed to enable faster movement in CVEs (e.g. glancing left and right). They also allow parallelism of actions with head movement and simultaneous two-handed interaction. Users can grasp one object, point at another and look around. At the same time, HMDs are fragile, not very comfortable and they have a limited field of view, but in the opinion of Hindmarsh et al. [2000], the latter is supposed to be compensated by the ability to quickly glance around. Projection-based systems consisting of one or more back-projection screens enable a wider field of view and enable the user to move more freely. Another option to increase the field of view is to use peripheral lenses which transform or distort the scene to enable a wider field of view. But several scenarios simulated in CVEs require that users have a clear view of the virtual objects without distortions, especially in automotive design reviews. The second recommendation made by the authors focuses on providing more information about other's actions. In common

CVEs, user's actions are represented by the source embodiment alone (e.g. raise an arm to point). But since source embodiment and target object are seldom in the same view, the action itself should be made explicit by additional visual information connecting source and target (e.g. rays or colored objects). In order to find suitable representations of actions in a CVE, these steps should be followed:

- Identify all necessary actions such as looking, speaking, pointing and grasping.
- Identify the targets of actions, wherein some objects are graspable and others could be referred to by speech or pointing.
- Determine how each relevant action is represented on the source, on the target and in the intervening environment, all of which should be consistent and distinguishable.

Regarding the representation of actions Hindmarsh et al. [2000] implement visible view frustra, speech bubbles, highlighting of objects and visible rays of light. In the thesis at hand, such artificial yet helpful cues are not implemented since the automotive users should find conditions as in their daily work with real prototypes, in order to not confuse and distract them. Obviously humanoid embodiments influence user perception in CVEs. But the effect of embodiments, however detailed they are, is questionable until users are provided with other capabilities of physical human bodies, such as a realistic field of view, gaze and body movement. Nevertheless, more exaggerated embodiments are required to compensate for limited views and possible behaviours within CVEs.

2.3 Social Human Communication

Various studies on supporting social human communication in collaborative virtual environments cover verbal and nonverbal communication as well as references to objects and the environment as introduced by Roberts et al. [2004], Corradini and Cohen [2002] and Otto and Roberts [2003].

Roberts et al. [2004] conduct an experiment using distributed walk-in displays in two different countries connected by a network. Linking a collaborative virtual environment with walk-in displays situates users in an intuitively social context. Such a technology supports the four primary elements of social human communication: verbal and nonverbal communication, references to objects and references to the environment. The goal of the study was to identify for each element of social human communication to what degree it is used in a collaborative scenario and to describe the requirements on consistency. Verbal communication includes speech and sounds, whereas nonverbal communication includes body language, such as postures and gestures. These elements are often synchronized as well. For example, one user points at a particular part while saying "lets get that". In this case, verbal and nonverbal communication are directly related. In walk-in displays it is possible to move around by natural body movement and to interact with other users and objects. The wide field of view enables natural head and body movement for both focused and general observation. In the experiment, avatars were used to represent the users' bodies but only the head and dominant hand were tracked. The scenario can be subdivided into four phases of collaboration:

- **Planning and Instruction** - Everybody should see and hear the discussion. To describe the task and to plan further steps, verbal communication is extensively used. Gestures like turning, pointing and nodding are used to emphasize the verbal communication. The more realistic the avatar is, the better collaboration is supported.
- **Working Separately** - In such phases, communication is reduced due to independence and the concentration on one's own task. Work related verbal communication is replaced by small talk and enhances the feeling of co-presence. In some cases, users offered assistance when the other one needed help. The wide field of view allows to stay in eye contact at the very least.
- **Moving an Object Together** - When moving an object together to a certain place in the environment, users have to know their responsibilities and actions. In the beginning, they must agree where they are going and how they pick up the object. On their way they have to adjust speed and communicate changes of path. Typically users try to keep each other in sight as long as the object is picked up (Figure 2.4).
- **Assembling Objects Cooperatively** - It is often necessary that one user holds an object in place while the other has to fix it with a certain tool. The single steps of fixation are verbally communicated. For example, if one user has to fetch a missing tool, the other should know to wait for him and keep the object in position.

Overall the use of walk-in displays in the experiment of Roberts et al. [2004] significantly improved the collaborative performance of the two users compared to a desktop system or an asymmetric combination of both display types.

Figure 2.4: Two remote users concurrently carrying a wooden beam in the virtual gazebo task introduced by Roberts et al. [2004].

Otto and Roberts [2003] give an overview of factors influencing communication in a collaborative task. They differentiate between verbal and nonverbal communication as well as the role of objects and the role of the environment in communication. In their opinion, collaborative virtual environments offer the potential for social interaction between geographically distributed groups. Closely coupled interaction is still very difficult in current collaborative virtual environments. This is due to the remaining differences between virtual and real worlds, concerning representation, consistency and responsiveness. Their research describes how the primary forms of human communication in the real world map to those in the virtual. Speech between participants is essential and the lack of other cues, such as gestures and postures increases verbal communication to coordinate and complete a given task.

It would result in an unreasonably high effort to communicate the informational content of a simple dialog by using only gestures and postures. The potential of virtual reality is to enrich a collaborative task by adding other cues, non-existent in reality, which compensate for limited or even missing verbal communication in distributed configurations. A major part of interaction happens via nonverbal communication. Gestures, facial expression, eye gaze and so on are cues that help in face-to-face interactions. Compared to desktop applications, immersive displays like HMDs offer more cues if a motion tracking system matches the spatial data from hand and head trackers to a virtual embodiment. The tracking data provides information about where someone is looking or where his hand points to. Objects in communication can be person (e.g. clothing) or non-person related. When two users share an object, whether sequential or concurrent or hand it over, then they are interacting in a non-person related way. Concurrent object sharing is not easy be it technological or communicational, but it can result in an increased amount of verbal communication. Imagine the virtual gazebo task (Figure 2.4) with two users having different opinions of how to move the beam to a certain position. This can result in an intense discussion. An interesting observation made by Otto and Roberts [2003] is that verbal communication between geographically remote users did not happen very naturally even though the audio devices were hidden. Verbal communication first increased when the users were constantly aware of a communication device like a headset. This might be different in co-located collaborative scenarios where natural voice is used indicating the presence and position of the other users (cf. Chapter 5).

In the experiment on pointing at virtual objects with real hands in Chapter 4, verbal communication between subjects was artificially excluded as suggested by Bangerter and Oppenheimer [2006]. Furthermore, the experimental virtual scenario does not make use of a surrounding real environment. The focus is on nonverbal communication and referencing objects by pointing gestures of the real hand which is in contrast to studies on virtual pointers by Chastine et al. [2008].

Chastine et al. [2008] determine the ability of users to generate and interpret reference cues as a critical component of successful collaboration. They research collaborative augmented reality wherein users need to refer to physical as well as virtual objects that surround them. A virtual pointer can be seen as the most basic referencing technique that satisfies these requirements. In a study, they explored the dynamics of group interaction and how virtual referencing methods can support collaborative tasks. Chastine et al. [2008] state that it is helpful for users to align themselves when giving and receiving references. Furthermore, several depth cues play an important role when interpreting references. In their experiment, an arrow with changing visual appearance pointed at different boxes which should be identified by the participants. In one condition, the workspace with the boxes was moveable and in another condition, only head movement was allowed. Intuitively, all participants tried to line up with the virtual arrow, which was easier with the workspace moveable than with extensive head movements. In a second experiment on the performance of pairs, the participants had to take over the role of guide or builder. The guide had to instruct the builder where to put certain objects by referencing with the arrow. At the beginning, the pairs made heavy use of referencing techniques to define an initial reference point in the workspace. Referential chains or the last referenced point as a relative basis were frequently used. It can be summarized that the inclusion of shadows enables more accurate references, and references depend on the configuration of the environment. It is suggested to provide shared viewpoints (e.g. shared video channels), allowing the users to switch views in order to avoid that one user is forced into a less desirable position. Virtual techniques are of great importance when creating an initial point of reference like a simple virtual object. Once this has been accomplished, other methods of referencing can be used as well, such as verbal communication.

2.4 Collaborative Interaction

Margery et al. [1999] defined three levels of cooperation. In their scenarios, two users coexist in a collaborative virtual environment and can perceive each other and communicate with each other (Level 1). Each user can individually modify the scene (Level 2) and simultaneously manipulate the same object (Level 3). The third level covers actions of users on a single object that are either independent (one user moves the object while the other changes its color) or codependent.

According to Ruddle et al. [2002], multi-user object manipulation is a field of cooperation that has rarely been studied. In an experiment they investigate object manipulation by pairs of users with the help of the piano movers' problem (e.g. maneuvering a large object through a restricted space). In their opinion, studies on cooperative manipulation have particular relevance to design reviews, data exploration and simulation and training. Especially within simulation and training, virtual environments are used to mimic real-world operations. Car designers could therefore investigate ergonomic problems of a new model by becoming virtual humans and simulating the installation of a dashboard step by step.

Ruddle et al. [2002] describe some important aspects that should be taken into account when designing symmetric interactions. It is necessary to define rules that manage inputs from the participating users (e.g. transferring the forces from each user's hands into object movement). Pure virtual object manipulation without haptic feedback requires dealing with physically impossible situations such as hand-object penetrations. Another important point is that if users are carrying a virtual object together they are not physically linked by the object. The same problem exists if a single user tries to manipulate a virtual object by using both hands. Since two users will not be able to move a virtual object in exactly the same way, a possible solution is to calculate the average of both user's input forces to compute the resulting movement of the object as suggested by Ruddle et al. [2002]. Another approach introduced by Pinho et al. [2002] is to give users the ability to manipulate one object at the same time by using different interaction techniques. For example, one user controls rotations while the other controls translations of the object.

Pinho et al. [2002] also state that new interaction techniques and tools need to be developed that enable more than one user to manipulate one object at the same time. They suggest a set of rules that define how different interaction techniques can be combined to achieve this goal. In most research work, cooperative interaction is not possible. If one user acts on a particular object this object is not accessible by others. Some approaches implementing collaborative work make use of force feedback devices to constrain a user's hand with forces generated by another user. Pinho et al. [2002] developed a software framework that allows combining different interaction techniques. They performed a study with pairs of users who had to place objects at certain positions or move a couch through a door. One of their questions was if collaborative manipulation leads to greater efficiency or ease of use in comparison to single-user or sequential manipulation. Two of their conclusions were:

- Cooperative techniques can provide increased performance and usability in difficult manipulation scenarios. However, single-user manipulation is simpler to use for most manipulation tasks.

- The use of a cooperative technique is applicable to those situations in which cooperation allows the users to better control some degrees of freedom (DOF) that cannot be easily controlled with the single-user technique.

When emulating real-world tasks in VR, it is often inevitable to provide cooperative techniques beyond single-user interaction. Collaboration between two users is often more complex; therefore, interaction techniques need to be even more comfortable and easy to understand.

2.5 Prop-based Interaction

Hinckley et al. [1994] show that the use of familiar real-world objects as a tangible interface may facilitate interaction in the virtual world. They used a tracked doll head or rubber ball to represent a tomography dataset of the human head (Figure 2.5). These passive interface props help users to understand the functionality of the represented objects or tools. Novice users directly interact with props and their corresponding virtual representation without requiring much training. The success of Nintendo's Wii indicates that the use of tracked handles connected to a virtual object is a strong interaction metaphor that may also partially compensate for missing force feedback.

The handling of props can be of great benefit in automotive virtual applications as well. Automotive engineers are employed to evaluate physical prototypes and handle real-world tools proficiently. In the thesis at hand (cf. Chapter 3), tracked handles are used to represent the handle of virtual assembly tools. The windshield prop introduced in Chapter 5 is an aluminum pole with a handle at each end. This two-user prop provides passive force-feedback and a direct physical link between the two users. It also represents the physical dimensions of the virtual object.

2.6 Seating Bucks

Seating bucks are a common tool used in the development process of many car manufacturers. Caputo et al. [2001] give an overview of the most common features of non-VR seating bucks and describe how they can be extended by VR-technology.

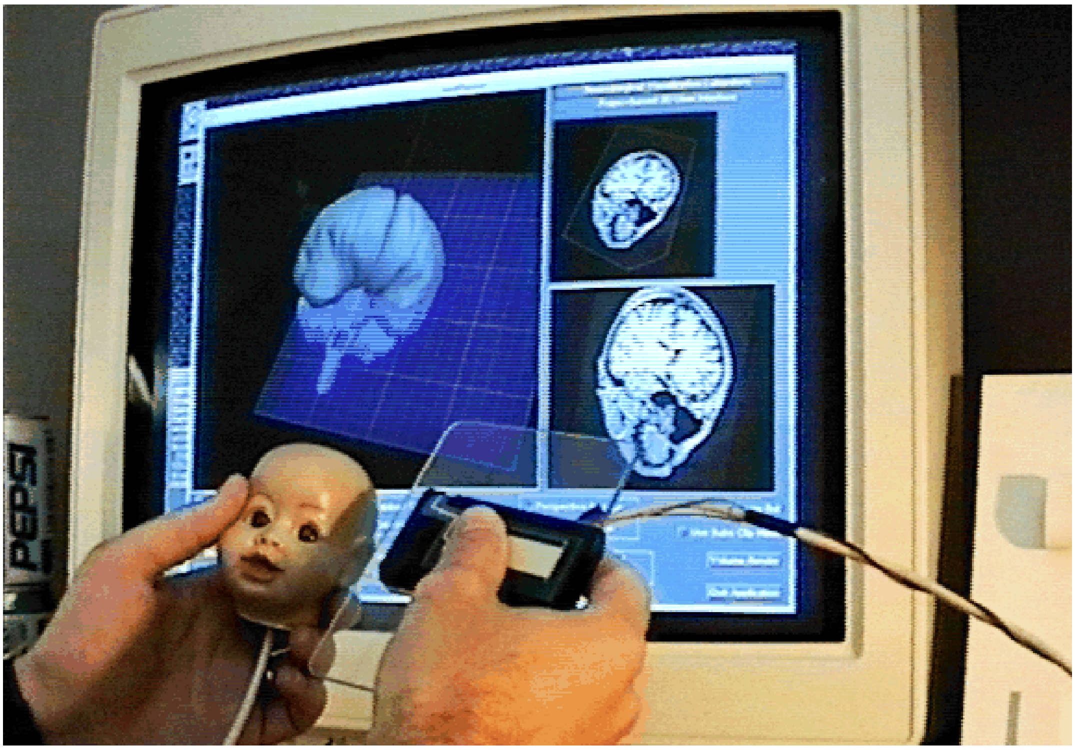

Figure 2.5: User moving a cutting plane through a three-dimensional image of the human head by using two props in a bimanual way as introduced by Hinckley et al. [1997].

It is also explained which steps have to be taken to build the virtual scenario (e.g. data preparation and calibration) and that a tracking system, besides aligning the virtual geometry with the real world, could additionally be used to extend such a virtual environment with virtual mannequins. In the automotive industry, VR-technology is used to enable designers, engineers and managers to evaluate different interior concepts in an early development stage and in this way improve the design of a new car step by step. A suitable virtual environment for automotive scenarios should include realistic visualization, a tracking system, stereoscopic devices, haptic input devices, human body models and it should be reconfigurable. From an automotive customer's point of view, this all is about to make sure that the driver and other passengers can see and reach all of the controls, understand the displayed information and feel comfortable in their seat. This results in a higher usability of the final product. In order to create a product focusing on humans of varying sizes, such a virtual environment helps to determine the performance of a human in a future car before physical prototypes have to be built.

Caputo et al. [2001] divide seating bucks in two groups: purely physical seating bucks, as they have been used for years in the automotive industry, and physical seating bucks extended by virtual reality technology. Purely physical seating bucks are configurable mechanical setups assembled out of basic elements for driver location (e.g. a seat or dashboard). They are reconfigurable but it is difficult and time consuming to adjust several components of a car to each other or to set up different variants. It is only possible to show one setup at a time. Seating bucks extended by virtual reality technology allow a lot more features to be displayed, including the following as a few examples:

- surround the user in the future car with a dynamic and consistent graphical representation of the interior
- enable and analyze user posture and movements to reach controls
- implement haptic feedback and retroactive forces
- explore several design alternatives in various configurations

Physical components of a virtual seating buck include at least a seat, steering wheel and foot rest, possibly with pedals, mounted on a platform that can be adjusted to physically represent different car types. The real components are visualization devices (e.g. HMD) as well as a tracking system to track head and hands. The tracking data is at first used to update images displayed in the HMD corresponding to user motion and secondly for posture evaluation with a virtual mannequin.

It should not be underestimated that several steps of virtual prototyping are necessary in advance before a virtual car with sufficient graphical quality is processed out of native CAD-data:

- collecting the model data including tesselation and decimation of polygons
- applying colors, materials, textures and lighting to the virtual model
- superimposition of virtual objects onto corresponding physical objects

Monacelli et al. [2004] state that the most advanced automotive companies, such as Volkswagen, have based their vehicle development process on digital mockups (DMU) and virtual simulations to reduce development time and costs. Virtual reality is thus used for styling, design, ergonomics, digital factory planning, marketing and sales. The DMU starts when the first part of a new car is designed in 3D and it is developed further and maintained over the whole lifecycle of a car model. Nowadays even the steps of the production process are simulated with DMU-tools. It is obvious that a large database of CAD-data is generated by representing each car model of a manufacturer and its creation with DMU.

Before DMUs were established, a lot of physical mockups (PMU) had to be built until the final model was perfected. Virtual reality offers the possibility to reduce the number of expensive PMUs by replacing them with virtual prototypes. In this way, bad or unattractive design alternatives can be discarded with the help of the virtual representation. The vision is that the first car produced is so advanced and mature that it could be sold.

Bordegoni et al. [2007] describe how mixed reality setups, such as virtual seating bucks, support virtual prototyping and what their benefits are. The main application area is the investigation of the ergonomics during car development (e.g. reachability of the light switch or the readability of information on the navigation system's display). They clearly identify car design as the industrial field where high costs of prototypes justify the investment in high-end virtual reality setups to enforce the implementation and acceptance of virtual prototyping. In their work, they use a setup including a see-through HMD, a motion tracking system, a human body model for ergonomic evaluations and a human machine interface (e.g. a knob and a display) to simulate real functionality in the virtual car and to provide haptic feedback. Since they are using a see-through HMD, a perfect superimposition of real and virtual environment is essential. So they address an important issue concerning the effects of HMD-optics, whether they are see-through or not, on the physiology of human vision. The questions are if absolute distances and depth are correct and how they can become more reliable. Inconsistencies may arise since the focus of the eye is fixed on its respective screen and the gaze of the user on the two screens is not tracked. Further possible sources of error include:

- distortions in the optics of the HMD
- mechanical misalignments in the HMD
- errors in head-tracking
- incorrect viewing parameters (e.g. field of view or interpupillary distance)
- end-to-end system latencies

Finally they propose to manually adjust selective subjective aspects affecting each other, related to the device and the user in order to reduce perspective and stereo problems.

Bordegoni et al. [2007] and Monacelli et al. [2004] employ a single user virtual seating buck while in the thesis at hand, this idea is extended to support the driver as well as the co-driver to enable face-to-face discussions of novel car interface concepts (cf. Chapter 6).

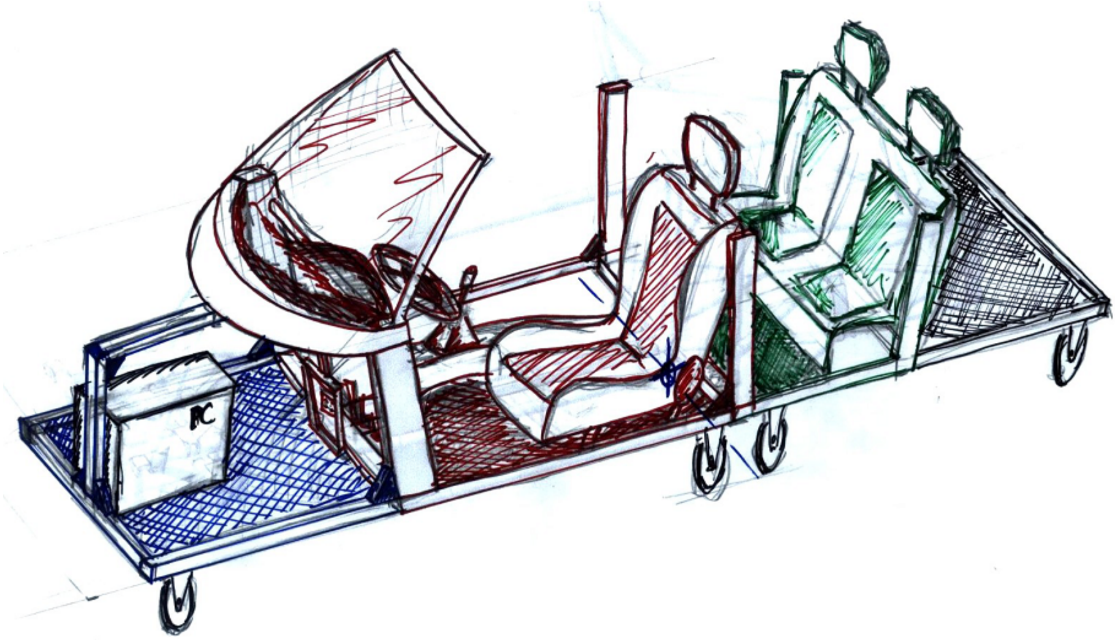

Figure 2.6: Sketch of a commonly used physical seating buck.

2.7 Virtual Avatars

Collaborative virtual environments support the interaction of multiple remote or co-located users in a shared virtual environment. In this thesis, usually two users are co-located in the same room and experience the same virtual scene. Tracked see-through HMDs or stereo glasses provide an individual stereoscopic view and enable free user movement as well as the perception of the surrounding real world, including other people. When using non see-through HMDs, it becomes inevitable to provide virtual body representations for each user as they are commonly used in networked virtual environments.

Slater and Usoh [1994] and Thalmann [2001] summarize the basic functions of these avatar representations: the visual embodiment of the user, the means of interaction with the world and the means of sensing various attributes of the world. The importance of avatar representations in HMD configurations should not be underestimated: Sanchez-Vives and Slater [2004] report that it can be even shocking for users if they are not able to see their own bodies. Thalmann [2001] further describes the crucial avatar functions for networked virtual environments, which also apply to co-located multi-user environments:

- perception (to see if anyone is around)
- localization (to see where the other person is)
- identification (to recognise the person)
- visualization of others' focus of interest (to see where the person's attention is directed)
- visualization of the other's actions (to see what the other person is doing and what is meant by gestures)
- social representation of self through decoration of the avatar (to know the other participants' task or status)

These roles of the avatar also apply to the scenarios in this thesis, wherein the visualization of the user's actions and focus of interest are particularly of relevance. The movements of these direct controlled virtual humans as stated by Capin et al. [1997] are based on information from the optical tracking system. Usually rather limited tracking information is available - head and hands only - nearly realistic body movement should be computed by inverse kinematics.

Jorissen et al. [2005] report on various studies on embodiment in collaborative virtual environments. In most of these studies, it is concluded that so-called virtual humans increase both the realism of interaction and the sensation of presence in collaborative virtual environments. According to Benford et al. [1995], the embodiments in this thesis can be classified as follows:

- **Presence** - the two users have more or less detailed humanoid looking embodiments with tracked heads and hands
- **Location** - each user has a relatively fixed location and limited orientation, corresponding to the respective tracking volume

- **Identity** - there are two different body models according to which role the user is playing
- **Viewpoints and Action Points** - the focus of interest for each user is represented by the orientation and position of his tracked hand and head representations
- **Gesture** - realized by finger-tracking gloves for each user
- **Facial Expression** - is not implemented, only static faces
- **Degree of Presence** - tracked hands/heads and voice

This list of issues is an initial guideline to successfully implement embodiments in CVEs. Each item should be considered relevant or not when designing an embodiment. Not all of these issues need to be accounted for since the level of detail of an embodiment mainly depends on the computing resources available. Furthermore, the relevance of each individual issue is application and user specific.

Chapter 3

Side-by-Side Assembly

In the automotive industry, assembly simulations are used for evaluating the ergonomics of an assembly sequence. In such scenarios the display of a full scale model and the involvement of multiple people are crucial for the validity and significance of these evaluations. For side-by-side interactions, two workers are typically standing in front of a car and are trying to mount different parts. To achieve this, the workers can employ different tools and common methods for navigation and manipulation. Generally, such simulations take place in front of a single projection screen where a group of engineers evaluates an assembly task by assuming the role of the respective workers (Figure 3.1).

All regularly used stereoscopic projection systems in the Volkswagen laboratories support only a single tracked user, which limits their usability for tasks involving two active users or involving a group of engineers. To improve the situation, a prototypical projection-based two-user system is set up providing two users with individual stereoscopic views. This setup is compared to and evaluated alongside the commonly used single-user projection setup in an expert review for collaborative side-by-side interactions.

This chapter starts with a short description of the used multi-viewer technology based on a projection system. For the following initial study, a two-viewer setup is implemented to investigate the advantage of a projection-based stereoscopic two-user system over a conventional single-user solution for common automotive scenarios and how it extends collaborative actions. In the next two paragraphs, typical automotive scenarios are described which are assembly planning and training. Based on these two scenarios, different expert groups identify advantages of the two-user setup in an expert review. The findings are summarized in the results section. Finally, this chapter ends with important observations and directions

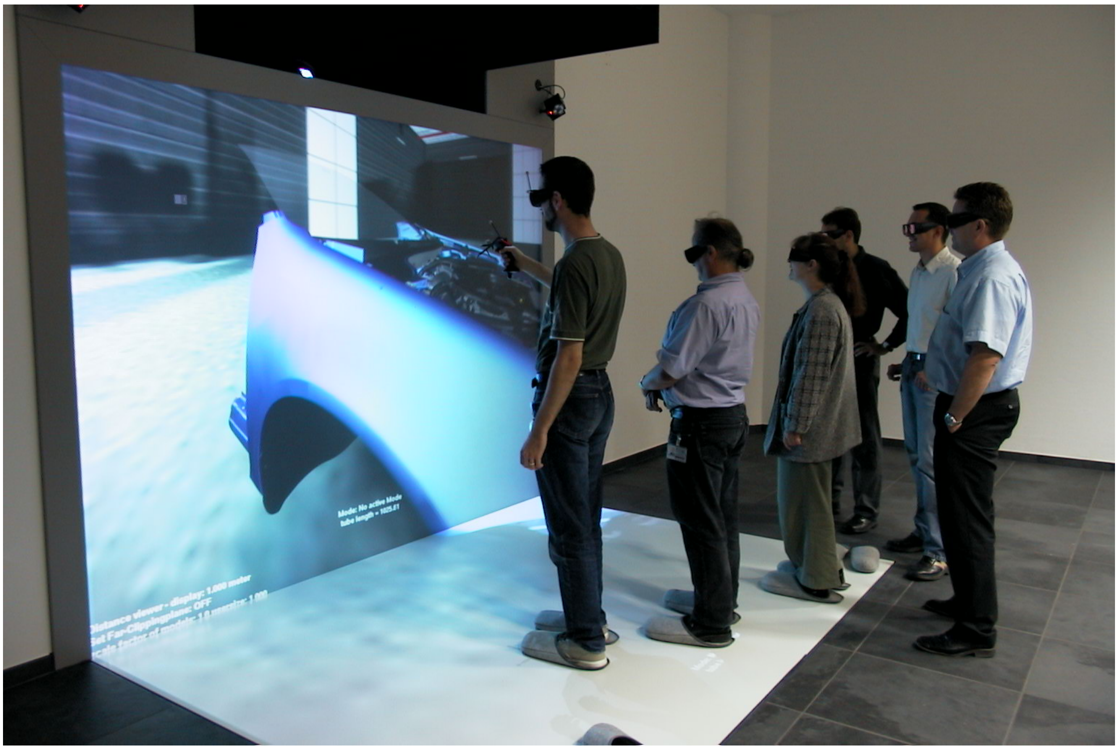

Figure 3.1: The far left user is the master with tracked stereo glasses and a tracked input device. The glasses of the other users surrounding him are not tracked. This leads to distortion of their views the further they are standing from the master.

for further experiments that were made by the experts during the evaluation. Regarding the aspects of collaboration defined in the introduction of this thesis, this part takes into account individual stereoscopic views, communication with others by natural voice and collaborative interaction supported by passive haptic feedback and real world props.

3.1 Projection-Based Two-User Setup

The two-viewer display system uses liquid crystal (LC) shutter elements mounted in front of LC-projectors to separate the individual users and provides a frame rate of 60Hz per eye per user as suggested by Froehlich et al. [2005]. Polarization is used to separate the left and right eye view of each user. The size of the display is 3 meters by 2.25 meters, the display's resolution is 1400x1050 pixels and the projectors' brightness is 3500 lumens. A high brightness is essential since the shutters limit the amount of light finally reaching the projection screen. The shutters in front of the projectors and wired shutter glasses are controlled by a custom micro-controller circuit (Figure 3.2). Each projector pair for a single user is driven by a separate computer equipped with a dual head graphics card (Figure 3.3). Both systems are driven by an in-house software, which keeps the states of both applications in sync. Standard LC-projectors emit already polarized light, which helps to set up such a system. However, the green channel is typically polarized orthogonally to the red and blue channel. For the left eye, the polarization of the green channel is rotated 90 degrees by the half wave plate, and for the right eye, the red and blue channels are also rotated 90 degrees. Thus the polarization of all three color channels for the left and right eyes are orthogonal to each other. Shutters consist of another half wave retarder embedded between two orthogonal polarization filters. Thus polarization is preserved and rotated 90 degrees.

The four images on the projection screen should be aligned perfectly, but at least the two images for one user have to be congruent to achieve a correct stereo vision. Otherwise the stereo perception will be disturbed.

3.2 Initial Study

The intention of this study is to investigate the advantage of a projection-based stereoscopic two-user system over a conventional single-user solution for common automotive scenarios. The first scenario is an authentic assembly planning task that was simulated some time ago in the development process of a new car. The original six participants, who are real experts in the field, were invited to be the interviewees. The focus was on the investigation of the accessibility of certain car parts by tools. The second scenario involves a typical video-based training of a new assembly method by a hands-on training sequence in a multi-user system. This extended training scenario was presented to five experienced automotive coaches in the interviews. What both scenarios have in common is that they require the execution of certain sequential interaction tasks as well as asymmetric interaction. Sequential means that they have to do similar actions using a single tool. For example, both have to fix screws, but there is only one screwdriver available. Asymmetric actions are those wherein one user moves the object and the other has to fix some screws. This way they are assuming different roles in the assembly task.

The two scenarios support similar interaction techniques. Users were able to point

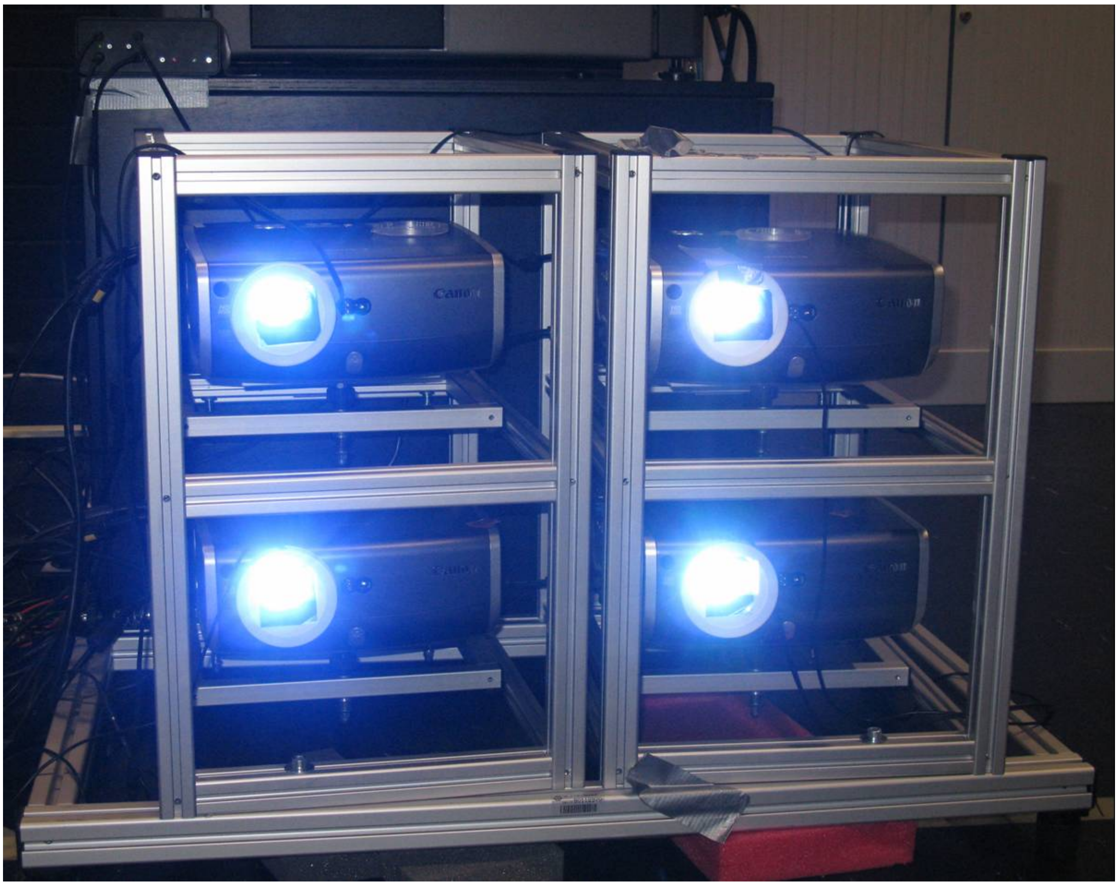

Figure 3.2: The setup consists of four shuttered projectors; one column for each user. The projector shutters are synchronized with the two corresponding shutter-glasses on the other side of the back-projection screen.

at different parts with their real fingers, which was enabled by the perspectively correct views in the multi-viewer projection system. Earlier tests indicated that pointing with the real hand is an intuitive interaction metaphor which is also highly appreciated in collaborative virtual scenarios. In the current experiment, two input devices (optically tracked handles with buttons) were provided that were used in two ways. One option was to have one device for the scene movement and the other controlled a virtual tool attached to it. This enables one user to navigate the scene while the other can check mounting points with the virtual tool in his hand. This task division was supposed to initiate communication between users. The user with the tool has to ask for viewpoint changes if he cannot properly see or reach the part he is supposed to manipulate. Alternatively a different tool was attached to each input device (Figure 3.4).

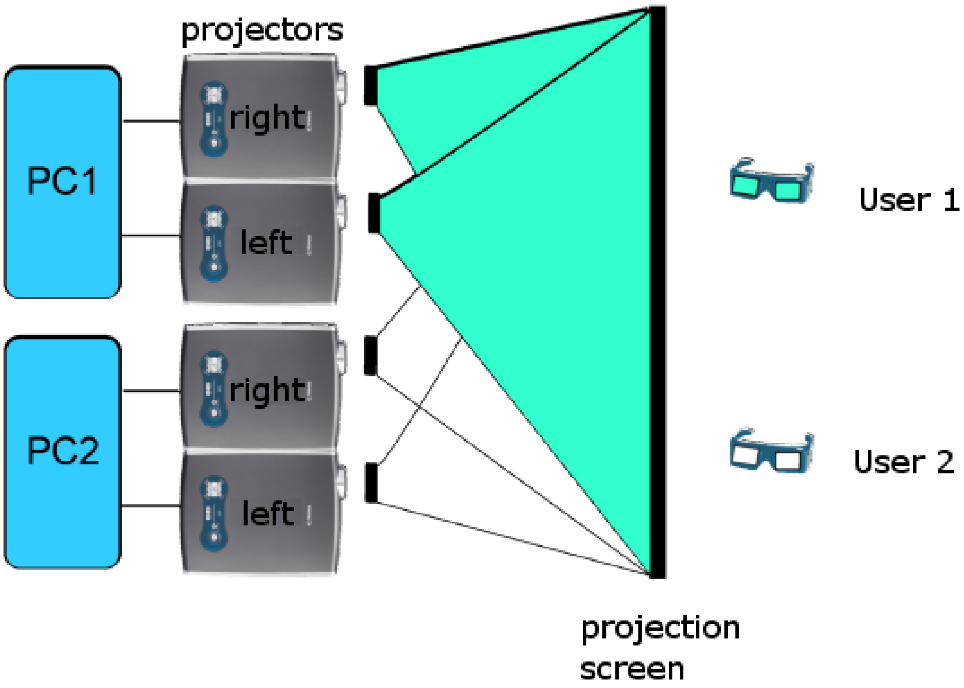

Figure 3.3: Projection-based two-user system. One user's eyes are separated by polarization while the users are separated by the shutters. This is enabled by a device mounted in front of each projector's lens. (adopted from Froehlich et al. [2005])

Since there was no haptic feedback evident, indicating collisions between car part and tool, a visual feedback was implemented. In the case of collision, the impacted polygons of part and tool were colored in red. To increase the ease of use when fixing screws, a snapping functionality was implemented. As soon as a tool entered a certain area of tolerance with respect to a screw, the tool automatically snapped onto the screw and rotation was constrained to one axis. In that way the following rotational movement was easier to perform because the tool had a more stable hold on the screw. To loosen the connection between both objects, the users had to move the tool a little more than natural. Regarding the virtual tools, an offset between the tracked position of the physical device and the position of the corresponding virtual tool should prevent the users from false or disturbed stereo perception due to superimposition. The intention of these implementations was to enable the participants of the study to concentrate on the collaborative task and not get lost or annoyed in too detailed operations.

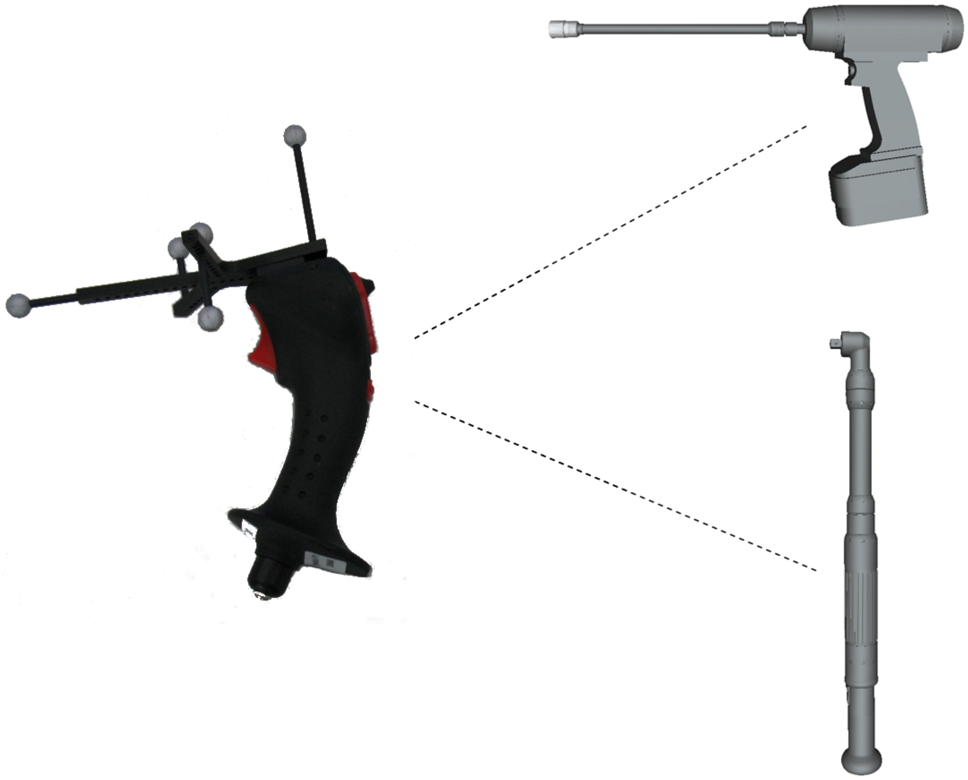

Figure 3.4: Two items of this tracked input device on the left were used as props for different tools. The physical device could be grabbed in an identical way as the real tool would be grabbed.

3.2.1 Assembly Planning Scenario

The first scenario is inspired by an existing protocol of a typical assembly simulation. The task was aimed to inspect different areas in the front area of a car. One part of the task was to fix the screw mountings of the car's hood which is connected to the car body. Two different tools are needed to fix the respective screws. In detail, the participants were asked to perform the following operations:

- The users initiate a discussion by pointing at certain parts with the real hand and naming the parts.
- One user checks screw mountings on the left side of the car's hood. The other on the right side. Since they are equipped with different tools attached to their input devices, they have to perform a handover of the used tool.
- One user has to move the front part of the car such that an otherwise occluded part becomes visible. The other user then has to fix some screws on this part

using a tool.

- The users have to navigate to see the windshield wiper and check for collisions on the surrounding car parts.

3.2.2 Training Scenario

The second scenario is a simulation of a training application. For certain assembly tasks, animated 3D-sequences are well accepted methods used to illustrate and train the mounting of car parts. Such a tutorial of the mounting of a part at the rear end of the car was displayed on a separate projection screen next to the two-user display (Figure 3.5). To improve upon the standard process, users were allowed to interactively perform the just-seen task in the stereoscopic two-user system. At the beginning, the users watched the animation sequence with the detailed process steps on the monoscopic display to become familiar with the upcoming task. In the second part, they performed the task immersed in the stereoscopic two-user projection. The scenario can be divided into these steps:

- The users initiate a discussion by pointing at certain parts with the real hand and naming the parts.
- One user has to fix screws of the back bumper with a tool attached to his input device both on the left and right back of the car. The other has to position the car in the desired orientation with the other input device.
- The users have to check several mounting points if they are reachable and fixable with the automatic screwdriver.
- The users have to exchange the input devices by handing it over to their partner.

3.2.3 Evaluation

For an initial evaluation of the usability of the two-user display, for both scenarios an expert review was performed involving automotive experts. User studies as defined in the literature are well accepted to evaluate immersive user interfaces, but they require a lot of resources and time. The research on human computer interaction has successfully implemented other evaluation methods such as focus groups, field studies and the expert review that apply more to the situation during

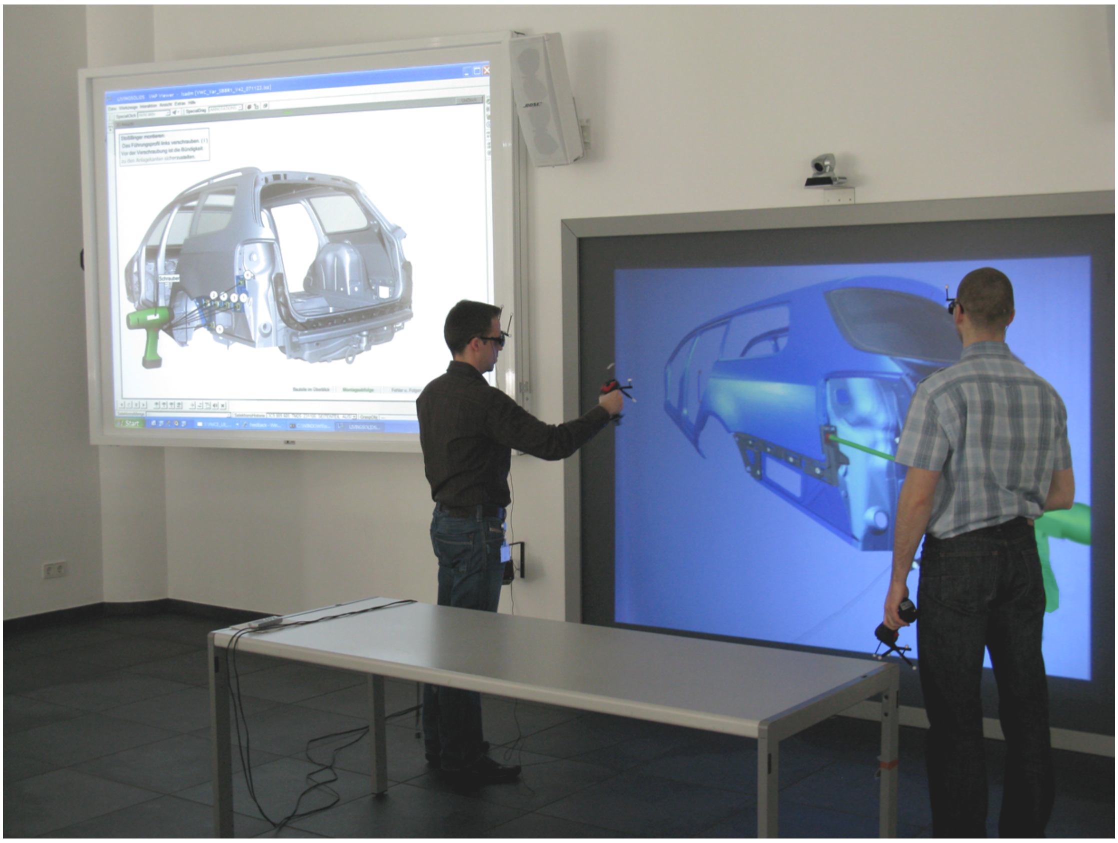

Figure 3.5: The training scenario allows users to first watch a video of the assembly sequence on a regular monoscopic display on the left, and then perform this sequence in the stereoscopic two-user system in single-user as well as in two-user mode.

automotive development processes. To explore the usability of the two-user system for the described tasks, an expert review was performed involving automotive experts. With this method as described by Preece et al. [2002], usually 75 percent of the usability problems can be found by asking only five experts. If otherwise only three experts are asked, 63 percent of the usability problems will be found. If the number of interviewees is doubled, this would result in only 12.5 percent more identified problems. Although this method is mainly used for usability evaluations of software and web interfaces, it is also applicable in VR-research for early design reviews of new system setups or interaction devices, as shown by Geiger and Rattay [2008].

According to Nielsen [1994] the results of the evaluation can be recorded either as written reports from each evaluator or by having the evaluators verbalize their comments to an observer as they go through the interface.

Written reports have the advantage of presenting a formal record of the evaluation, but require an additional effort by the evaluators and the need to be read and aggregated by an evaluation manager. Using an observer adds to the overhead of each evaluation session, but reduces the workload on the evaluators. Also, the results of the evaluation are available fairly soon after the last evaluation session since the observer only needs to understand and organize one set of personal notes, not a set of reports written by others. Furthermore, the observer can assist the evaluators in operating the interface in case of problems, such as an unstable prototype, and help if the evaluators have limited domain expertise and need to have certain aspects of the interface explained. This is an important feature regarding the evaluation of the two-user setup with not commonly used input devices.

In principle, the evaluators decide on their own how they want to proceed with evaluating the interface. However, a general recommendation would be that they go through the interface at least twice. The first pass would be intended to get a feel for the flow of the interaction and the general scope of the system. The second pass then allows the evaluator to focus on specific interface elements while knowing how they fit into the whole system.

If the system is intended as a walk-up-and-use interface for the general population or if the evaluators are domain experts, it will be possible to let the evaluators use the system without further assistance. If the system is domain-dependent and the evaluators are inexperienced with respect to the domain of the system, it will be necessary to assist the evaluators to enable them to use the interface. One approach that has been applied successfully is to supply the evaluators with a typical usage scenario, listing the various steps a user would take to perform a sample set of realistic tasks. Such a scenario should be constructed on the basis of a task analysis of the actual users and their work in order to be as representative as possible of the eventual use of the system. In the current evaluation, this should be achieved by the described scenario steps of assembly and training scenario mentioned above.

One possibility for extending the heuristic evaluation method to provide some design advice is to conduct a debriefing session after the last evaluation session. The participants in the debriefing should include the evaluators, any observer used during the evaluation sessions and representatives of the design team. The debriefing session would be conducted primarily in a brainstorming mode and would focus on discussions of possible design changes to address the major usability problems and general problematic aspects of the design. A debriefing is also a good opportunity for discussing the positive aspects of the design, since heuristic evaluation does not otherwise address this important issue.

As an evaluation method, a semi-structured group interview was chosen, allowing new questions to be brought up which result from the discussion with the interviewees. The interview is guided by general objectives that should be explored and it does not have to stick to a fixed set of questions. The general objectives are summarized in an interview guideline. During the interviews, the groups were accompanied by one interview conductor and one person noting the comments made by the group members. The conductor led the interview loosely by asking questions regarding the general objectives from time to time. The interviewees were able to comment on the scenarios at any time. Each interview took about 20 minutes.

As mentioned above, the assembly planning scenario was evaluated by six automotive experts who were actually engineers. The training scenario was presented to five experienced automotive coaches. A third group was formed out of five VR-researchers who evaluated both scenarios. With this composition of groups, it was intended to receive results that cover task specific issues (from engineers and coaches) as well as technical issues with respect to the VR-technology (from VR-researchers) used. The coaches had the lowest level of experience with VR-technology followed by the engineers with a medium level of experience, and the VR-researchers had the highest level. With the help of the general objectives, it was possible to generalize the comments of the participants and to formulate quintessences between groups. If misunderstandings arose, the interviewees were asked again to verify the respective comment.

In the beginning, the participants were instructed that their task was to compare the new two-user setup to the usual single-user setup. A group of interviewees had to go through the introduced process steps of the respective scenario. For each scenario, users first performed their tasks in single-user mode, wherein only one user was tracked and the second user saw the images computed for the first user's viewpoint. Then both users were tracked and provided with individual stereoscopic images (two-user mode). Finally, users were exposed to single-user mode again. Thus each pair of users performed a scenario three times, which took approximately 25 minutes. The first pass allowed the participants to become familiarized with the scenario task. In the following passes, they were then able to compare both configurations.

3.2.4 Results

Due to the general objectives included in the interview guideline, it was possible to find quintessences that were made in each single group as well as between the groups (Table 3.1). They give a first indication of the results while considering only a small part that was equally and independently stated by the participants of all groups. A more detailed analysis of statements made in the interviews follows in the next passages.

Table 3.1: Quintessences grouped by general objectives and the respective agreement of the interviewees.

general objective	**particular item**	**agreement in percent**
interaction	handover of tools better in two-user mode	100 %
	pointing better in two-user mode	60 %
comfort	wiring of glasses is disturbing	100 %
	glasses slip off the head too easily	60 %
presentation	virtual hand is not missed	100 %
	virtual avatar is not missed	100 %
improvement	design of glasses could be better	100 %
	projection screen could be bigger	40 %

Independent of the scenario, the experts stated that it is very hard or even impossible for the single-user slave (non-tracked user) to interact in any way since the single-user master's (tracked user) head movement affects his view. Some of them had the fear of becoming sick, considering that the movement was not controlled by them and occurred without their complete awareness. They also realized that the artificial single-user mode has the same shortcomings as the stereoscopic projection system they were used to, and that the distortion gets worse the further away they are from the tracked user. Some of the participants mentioned that in earlier immersive presentations, novice users from other departments were not told that they had a distorted view and so they accepted the distortions as a given disadvantage of virtual reality technology overall.

The asymmetric task division, wherein one user controlled navigation and the other a tool, was not well received. In particular, training experts said that it would be a better option to give one experienced user of virtual reality technology (e.g. the teacher or instructor) the ability to control navigation and menu functions to avoid a chaotic presentation in a potentially larger group of users. In a real assembly situation, there is no need to navigate the scene because the observers are usually moving around the car. In the current setup this function was important to compensate for the relatively small viewing volume which prevents the users from walking around the car as they would in reality. The possibility to hand over a tool or to exchange tools with different functions between users was seen as an improvement compared to single-user mode. Users often switch their roles by simply exchanging input devices. Figure 3.6 shows the handover of a tool by passing the input device to the other user. For the single-user slave, the tools had the same distortion as the parts of the scene, which made a correct grasping and operation impossible. Another effect was that distorted parts and tools were clipped at the boundaries of the viewing frustum.

The effect of clipped objects at the boundaries of the viewing frustum can be partly compensated by one user controlling the scene movement, but the interaction space is clearly limited by the single projection screen. The interviewed experts confirmed that it is unnecessary or even confusing to have virtual hand representations since users can see the hand gestures of their counterpart and their own through the stereo glasses. Virtual hand representations trigger the users to switch their focus between the real and the virtual hand, which disturbs the stereo perception. These findings are in contrast to HMD-scenarios wherein virtual representations of hands and heads are essential since the view on real body parts is blocked by a non-see-through HMD.

The expert-interview has shown that in a setup consisting of only a single projection screen, it is not essential to have avatars representing the users or their hands. In most cases, the users are standing relatively close together and do not look at each other. In addition, shutter glasses allow the users to see their real bodies and hands such that gestures and postures can be well communicated. In both scenarios, no virtual hand was used during the interviews since it could be expected that it is not that important in a projection-based setup where tracked

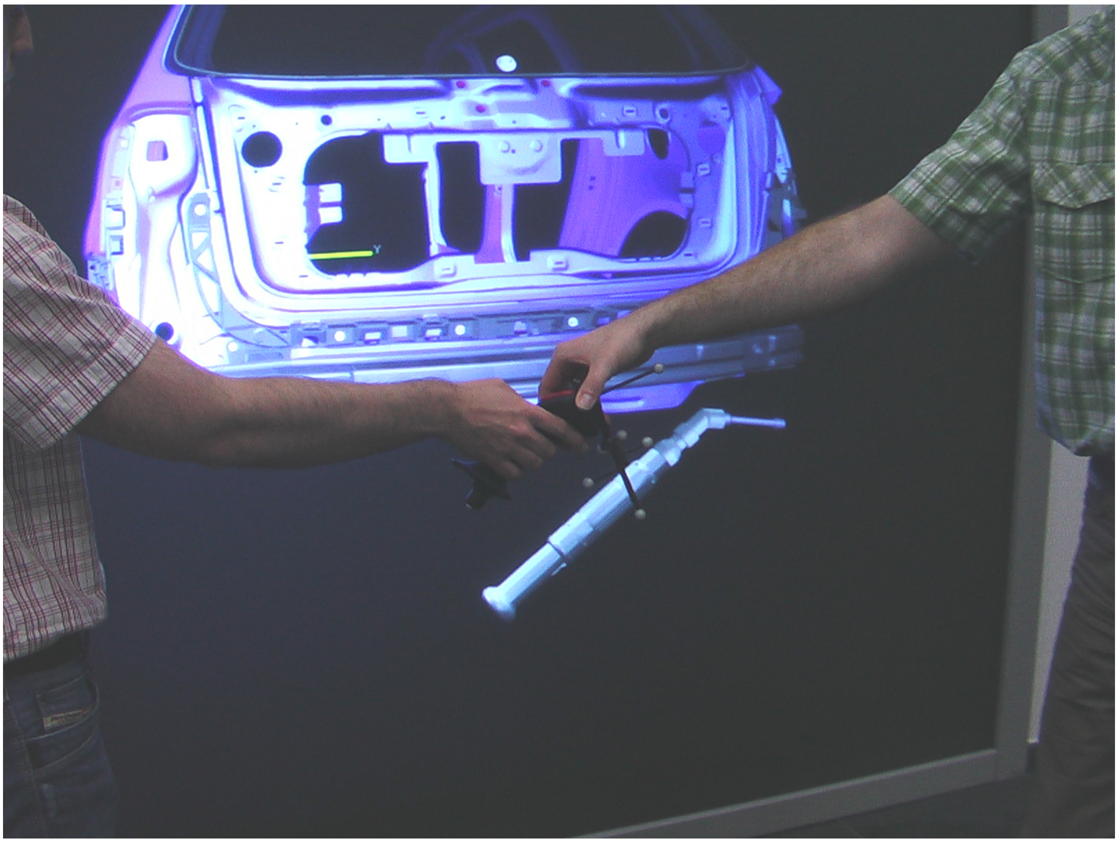

Figure 3.6: The users are able to hand over a virtual tool that is mapped onto a tracked input device while investigating several screw mountings.

handles are used as props for graspable tools. All of the experts confirmed that they did not miss their virtual hand. One of them said that a virtual hand is only an additional object that occludes the more important virtual tools and car parts.

The experts suggested integrating more users with different roles. An interesting point is to have two or more users looking from different directions into the car or users taking different positions related to the car and interacting collaboratively. For example, one in and one out of the car or one left and one right of the car, would enable more effective and faster working. But if the users are standing in front of one interaction/projection plane in reality, it is possibly hard to understand that they have different positions in the virtual environment than in the real. So the possibilities depend on the projection screen configuration used. Another option suggested was to group the users by departments according to their expertise. For example, a group of car body designers and a group of electricians. Each group has a designated master with tracked glasses and the remaining members of that group have non-tracked glasses synced to their master's glasses. So the non-tracked users of one group would still have a distorted view, but it would be a simple way to increase the number of participants in the current setup. However, more independently tracked viewers with individual stereoscopic views were clearly identified as the better option.

Overall, the results with the initial prototype indicate that interaction support in such co-located collaborative environments need much more research to enable real-life like experience. During the expert interviews, quite a number of relevant observations were made.

3.2.5 Observations

The participants of the interviews also suggested improving some hardware issues. For example, the shuttered stereo glasses could be improved in some ways. The actual glass in front of each user's eye proved too small in order to provide a complete stereoscopic field of view. Furthermore, it was suggested to block the view left and right of the glasses' frame with some sort of blinders. The glasses were wired and so it often occurred that they almost slipped off the user's head. This fact has a direct impact on user comfort and should be improved. Regarding the projection of the four images, for each user's eye, sometimes ghosting effects became apparent. This indicates that the shuttering is not sharp enough or that the polarizers in front of projector lens and corresponding eye glass are not perfectly aligned. Some of the participants wished for a larger projection screen to visualize 1:1 car models and to prevent clipped objects at the boundaries of the projection.

Grade of Realism - A major point one has to consider when designing assembly simulations is if the virtual simulation should be as realistic as possible (e.g. physically correct, multi-hand grabbing, etc.) or if the original advantages of virtual reality (e.g. objects with no weight, ray selection, etc.) count more. The desired degree of realism mainly depends on the nature of the task. For example, if one wants to simulate assembly paths in an early development state of a new car, this could be solved with less realistic virtual interactions. If the task is more to simulate maintenance tasks, the interviews showed that more realism with multi-hand grabbing and gravitation forces are essential. Since service work and repairs are usually done by hands of multiple users and not with robots, like at the assembly lines, maintenance planners have a big interest in using a multi-user system for simulating maintenance tasks in a realistic way.

Multi-Hand Grabbing - To increase the realism of interactions in a virtual assembly simulation with multiple users, it is inevitable to support multi-hand grabbing in one of the following options:

- 1 user grabbing one object with two hands
- 1 user grabbing two objects one in each hand
- 2 or more users grabbing one large/heavy object
- 2 or more users grabbing different objects

If this grade of realism is desired, a VR-software used for simulations with more than one user has to support scenegraph-syncing. Furthermore, correct physical behaviour of objects and different options of object-ownership have to be taken into account.

Manipulating Large Objects - During the work with the two-user projection system interactions with large objects in a car were also tested, with the windshield or the roof interior as examples. The current setup of the system allows two users to work collaboratively in one interaction plane. When mounting a virtual windshield in the car, one user is positioned left and the other right of the car, which enables them to grab the windshield on each side and simultaneously move it into the correct position. When mounting the roof interior in a car, usually one user is positioned inside of the car and the other outside so they have to synchronize their movements by verbal communication. The roof interior has to be grabbed on each end by the two users and is then moved through the windshield opening. That means both users have to look at each other while assembling, but this is not possible in a projection screen configuration as presented here. This example indicates that interaction with multiple users is depending mainly on the projection screen setup. To enable even more types of collaboration, it is necessary that users can take positions with viewpoints turned around 90 or 180 degrees relative to the other user's viewpoint.

Avatars - The expert interview has shown that in a planar setup with one projection screen, it is not essential to have avatars representing the users. The users are mostly standing relatively close together and do not look at each other. The shutter glasses allow the users to see their real bodies as well, as in most projection-based systems. So either the users are looking at each other and not in the direction of the projection anymore, or the users are looking at each other and in the direction of the projection. In the latter case, the body of the counterpart blocks the view on the projection screen. However, it does not make much sense to display virtual body parts since they will always be hidden by the real body. In both scenarios of the initial study, no virtual hand was used during the interviews. All of the experts confirmed that they did not miss their virtual hand. One of them even said that it is only an additional object that occludes the more important tools and car parts.

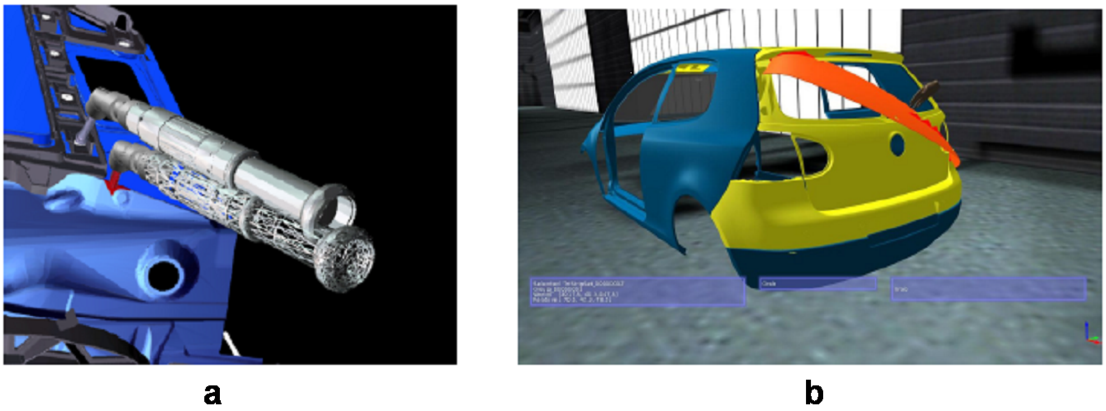

Figure 3.7: Two possible options of visualizing collisions in an assembly scenario. a) in case of collision the tool is displayed as a wireframe and frozen for a short time and b) the whole back window is colored as long as the collision persists.

Visualizing Collisions - Regarding the visualization of collisions, it was observed that especially for large objects, a coloring of the colliding object polygons is very disturbing for the users. On the other hand, it is not possible to locate the correct spot of collision when the whole colliding object is colored (Figure 3.7). A more preferred alternative, at least in automotive scenarios, is to have a gliding function which prevents the respective objects from penetration. In earlier tests it was observed that even with no force feedback-device, some users constrained the movement of their real hand when the virtual representations of tools or hands changed their movement due to the gliding function.

3.2.6 Conclusions

The initial experiment comparing a single-user projection system to a two-user setup showed that individually correct stereo perspectives greatly facilitate side-by-side interactions of two users such as pointing or using different tools.

To increase the realism of interactions in a virtual assembly simulation involving multiple users, it is inevitable to support multi-handed grasping. Two or more users grasping one large object or two or more users grasping different objects that should be assembled are such examples. A pseudo-physical or even a fully physical simulation of the virtual environment was seen as a major requirement to enable collaborative interactions. In addition, object-ownership has to be taken into account to avoid access conflicts to the objects if a physical simulation is not available.

In the expert interviews, it was also asked for future scenarios in a multi-user projection setup. One suggested task was the mounting of the windshield or the back window where both users have to grasp opposite sides of the window frame and they have to position it carefully. The suggested windshield assembly task will be reused in Chapter 5 in an evaluation on prop-based collaborative interaction. Another example is cable laying with one user inside of the car while the other helps from the outside. Their task is to attach a set of cables to the car body. A perfect application scenario is a common assembly task in the front of a car, where two workers are standing very close to each other between the front wheels. They have to perform several preparation steps before the motor can be placed.

3.3 Summary

All regularly used stereoscopic projection systems in the Volkswagen labs support only one tracked user, which limits their usability for tasks involving two active users. To improve the situation, a prototypical projection-based two-user system was set up providing two users with individual stereoscopic images. This setup was evaluated in an expert review for collaborative side-by-side interactions and was compared to the commonly used single-user projection setup. The results showed that the two-user system greatly facilitates basic collaborative interactions and it proves to be a standard evaluation platform for side-by-side tasks.

Chapter 4

Pointing as Basic Interaction

In commonly used projection-based setups, only one user of the group is wearing tracked stereo glasses. Due to the distortions for non-tracked viewers, the tracked glasses had to be passed around. This is very time consuming and the discussion process is disrupted so that at no point in the discussion can two users directly refer to a virtual object by simply pointing and talking about it. Thus, often virtual pointers are used and the users try to avoid moving around or getting too close to virtual objects.

Perspective projection in combination with head tracking is widely used in immersive virtual environments to support users with correct spatial perception of the virtual world. Recently, projection-based stereoscopic multi-viewer systems have become available as introduced by Froehlich et al. [2005], which provide individual stereoscopic images for multiple tracked users. In the context of virtual automotive simulations, experts use stereoscopic projection systems to find a solution for an engineering problem. They come together in a room and discuss the recent problem based on the content visualized on the stereoscopic projection screen. This display technology is well suited for the automotive industry wherein collaborative 3D-interaction of a group of experts is often desired. These expert groups can consist of electricians, assembly planners, ergonomists and others. In such scenarios, the large projection screen serves as a central communication platform.

One interesting aspect brought up during the expert review was whether pointing at virtual objects can be performed with the same precision in virtual environments as in reality. Pointing is a simple yet fundamental interaction in collaborative scenarios. In a single-user stereoscopic environment, discussions about assembly tasks can be very frustrating for the participants if the person with the correct view points with his bare finger at a virtual object. In this case, the finger points at a different position in the virtual environment for each of the other users because of the distorted perception of the virtual scenario. Users can verbally describe the object they are pointing at in more detail, but sometimes it occurs that not every other participant knows the specific technical expression used due to their different backgrounds. Another option would be the usage of finger tracking, but experts' time is limited, and time-consuming and obtrusive hardware equipment for finger tracking should be avoided. Due to the often tense atmosphere during such discussions, with people representing their respective field of expertise, it is important that the virtual techniques fit into the discussion process seamlessly and people can behave naturally.

The two-user setup has the potential to support correct pointing at virtual objects with the real finger for two users (Figure 4.1). Informal tests of the system already showed that pointing worked quite well overall and that participants had no mentionable problems when pointing with real fingers at virtual objects. However, this interaction metaphor could suffer from focus and convergence mismatch and occlusion effects.

In automotive simulations, one user often needs to communicate specific information about a particular part of the car to a second person. During the initial study with the expert review, users intuitively used their hands to point at a car part, which worked well in most cases. However, some experts were wondering how accurate pointing in virtual environments is in comparison to real-world pointing.

To answer this question, a study was designed to compare different pointing techniques. Those were evaluated in pointing tests for a number of different objects in virtual and real-world conditions. The results should indicate whether these techniques are more error prone in virtual environments than they are in the real world.

In this chapter, a basic interaction metaphor - pointing with the real hand - is investigated in more detail. During the initial study in Chapter 3, the experts intuitively pointed at virtual objects with their real hands. Due to this, the question arose if virtual pointing is as exact as pointing in the real world. At the beginning of the section, typical modes of pointing are identified and the real dashboard versus a virtual dashboard comparison task is introduced. This is followed by remarks regarding calibration, which is an important aspect when comparing pointing accuracy in an identical virtual and real scenario. Finally, the results are presented and recommendations for future scenarios supporting pointing with the real hand

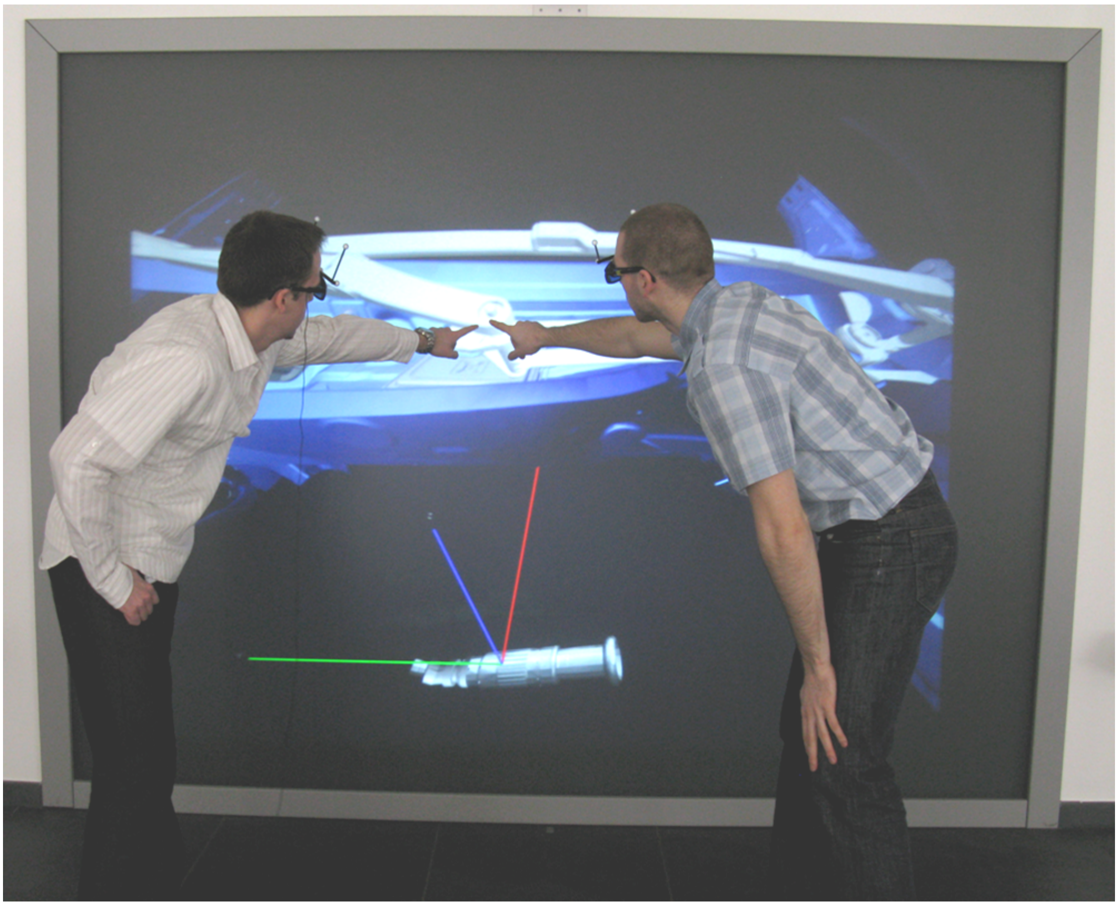

Figure 4.1: Users are discussing collisions of a windshield wiper joint in a two-user projection system for collaborative assembly simulations.

at virtual objects are given. The aspects of collaboration addressed in this part are individual stereoscopic views and the referencing of objects by pointing gestures.

4.1 Experiment Design

There are different ways how people identify an object by pointing when they try to communicate some information about that particular object to another person.

- **Touching** - In reality, it is very easy for observers to identify an object which is directly touched by someone. In projection-based display systems, one would assume that the stereoscopic perception is affected if the real hand is in proximity to a virtual object due to conflicting occlusion as well as focus and convergence cues. Thus, it should be hard to estimate if a real finger is in front of an object,

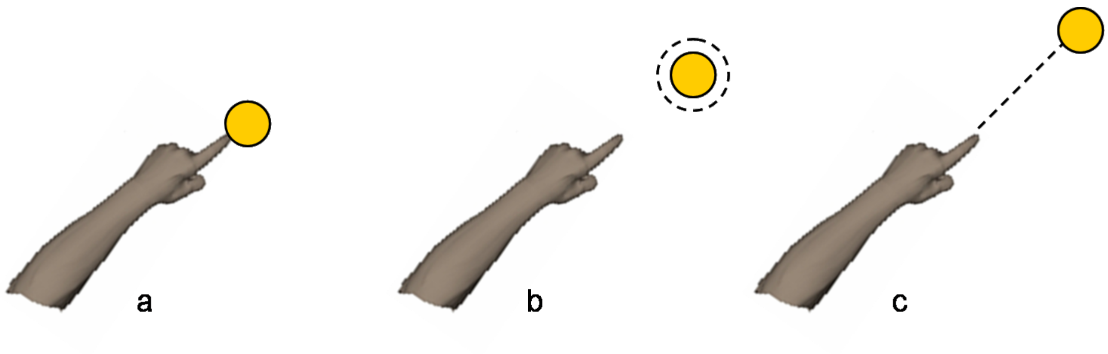

Figure 4.2: The three different pointing techniques. a) directly touching an object, b) drawing an outline around a distant object, c) pointing at an object from a distance.

inside or already behind.

- **Outlining** - It is slightly harder to identify real objects by drawing their outline in the air with little distance to the corresponding object. It becomes more difficult if more objects of the same shape and size are positioned close to each other but it can be assumed that the same effect occurs in virtual environments.
- **Distance Pointing** - The most difficult type of pointing in the real world and in virtual reality is pointing at objects from a distance, in particular if objects lie close to one another.

While there are certainly other possibilities to reference a particular object, the following is focused on these three selection techniques:

Figure 4.3: The right picture shows the pointing experiment in reality. One subject points at a given target object while the other has to decide where he is pointing. The left picture shows the corresponding virtual pointing condition.

In order to find a suitable object for the pointing task, several aspects had to be taken into account. A first idea was to reuse one of the objects from the two initial multi-user test scenarios, but there were too many technical expressions and names of parts that are not known to everyone. Furthermore, an object matching the projection setup, which supports only one projection screen, had to be found. In the two test scenarios it often occurred that virtual objects were clipped at the boundaries of the projection screen due to missing floor and side projections. The dashboard was identified as most suitable because everyone knows most of its elements by name and uses them in every day life. Some of them are even labelled. Furthermore, the dashboard matches the size of the projection screen and has objects at several depths. Another point is that it includes large objects as well as small objects with different shapes.

A real dashboard versus a virtual dashboard comparison task was set up. Two participants were asked to stand in front of the dashboard. One person pointing at objects; the other person guessing which object was referred to. This task was performed in front of the real and the virtual models. Each participant had to take over both roles: pointing and guessing. Each pointing participant had to point at three touch objects, three outline objects and three distance objects of the virtual and the real dashboards. The target objects were never repeated during one complete session and they were different for the real and virtual dashboards. That means one pair of participants had a set of eighteen objects from the three different classes (Figure 4.4). Orders of objects randomly changed between pairs of subjects, and half of the groups started with pointing at the virtual dashboard while the other half started with pointing at the real dashboard to avoid learning effects.

During the pointing task, subjects were not allowed to move except for arm and small head movements. An instructor showed the pointing participant the target objects he had to point at on a picture of the dashboard, not visible to the other participant. No spoken commands were uttered until the end of the task. Only the observing participant was allowed to say which target object he identified. The instructor recorded the matches and mismatches.

The evaluation method was partially adopted from previous studies. For example, Corradini and Cohen [2002] conducted an inspiring experiment on the precision of pointing. They let subjects point at certain targets on a wall from different distances with a turned off laser pointer. If the participants were sure to aim at the intended position, the laser pointer was switched on and the distance to the target

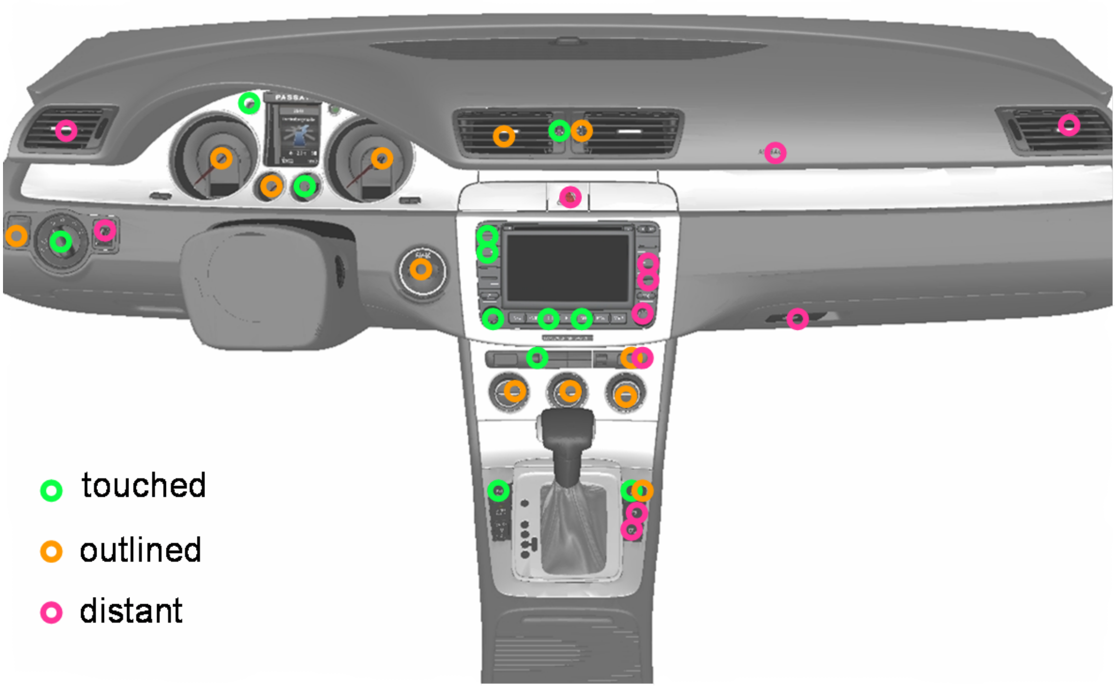

Figure 4.4: The allocation of the target objects corresponding to the three object classes.

on the wall was measured. To isolate possible sources of potential errors in pointing at virtual objects, the goal is to figure out with which accuracy the pointing subject was able to hit objects. This capability should be strongly impacted by focus and convergence mismatch and an error in pointing can be expected, as well as an error in perceiving finger positions by the observer. To measure accuracy, a small tracking target was mounted on the index finger of each pointing subject. When the subject pointed at the virtual objects that had to be directly touched with the tip of the index finger, the position was measured. This allows the possibility of calculating the distance between real fingertip and the touched virtual object. In the real scenario, this distance should be almost zero, but size of the finger and size of the object influence this measure as well.

Twenty-four male subjects aged 25 to 45 volunteered to participate in the study, all coming from different backgrounds including assembly planners, VR-researchers and students with different levels of VR-experience. They performed the pointing tasks in pairs of two and filled out a questionnaire afterwards. The questions aimed at demographic issues, VR-experiences and the pointing task itself. Most of the questions had to be answered on a 1 to 5 Likert-scale. Here are some examples:

- How often do you use [projection walls, CAVE's, HMD's, 3D-Games, 3D-Cinemas, driving simulators]?
- How much did you like to point at virtual objects with your real hand?
- How well could you estimate the position of your index fingertip in front of the projection wall?

4.2 Calibration

A high quality of the virtual environment setup was essential for the relevance of the study results. Projector matching and tracker calibration had to especially be done very carefully. Introducing errors in calibration of the virtual environment would have distorted the results of the evaluation. The virtual environment was defined by carefully measuring the hardware setup. Inter-pupillary distance (IPD) was not adjusted individually, but an average distance of 65 mm, as suggested by Dodgson [2004], was assumed. This also applies to daily automotive work, in which groups of people use the system and glasses are passed around in such groups. An individual adjustment of IPD for each user would mean a measuring procedure at the beginning of the virtual session and an adjustment procedure during the session when glasses are exchanged. The goal was to find out which accuracy can be expected from such a realistic use of the system environment. The projectors were aligned manually, introducing a slight mismatch at the image boundaries due to limited adjustment capabilities. In the middle of the projection screen the images were perfectly aligned. The smallest and most detailed objects for the pointing task (e.g. the buttons of the navigation device (Figure 4.6), were placed in the central screen area.

The calibration was tested before starting the evaluations by visually comparing both users' projection having the same camera pose and by visually checking registration of a tracked object. There was no eye-to-eye or user-to-user mismatch visually detectable in the screen area used for the evaluation. However, no detailed error analysis was performed. The real dashboard scenario was placed on a table right next to the projection. The virtual scenario was set up in a way that it appears at the same height and depth from the participants. The pointing participant took a standing position in front of the left part of the dashboard, with the majority of the objects in reach of the right arm, and the observing participant took a standing position in front of the dashboard's right side (Figure 4.3). Their distance to the dashboards was 50 cm. Both had a distance of 1.2 m to the wall of the room which corresponds to their distance to the projection wall.

4.3 Results

The pointing accuracy measurements for the interactions requiring touching an object are a good indication of how well a pointing participant could localize his finger in relation to a virtual object. Based on the statements in the questionnaire regarding VR-experience taken from the scaled questions mentioned above, three groups were created among the participants with respect to their use of virtual reality technology. They were labeled "often", "sometimes" or "seldom". The mean pointing accuracy was calculated for each group by computing the distance between each virtual object's face-center closest to the participant and his index-fingertip. Virtual object size and finger size should have been taken into account, but it was expected that the users tend to always point in the middle of the object. Furthermore, it is not known which point of his finger the pointing participant aligned with which point on the target object and his eye. In the virtual condition, there was also no shadow of the finger visible on a touched object.

Table 4.1: Means for the accuracy of matches between real index fingertip and virtual target object among the three subject groups for the condition that the object had to be touched.

vr usage	**pointing accuracy in cm**
often	2.36 ($\sigma = 1.68$)
sometimes	3.31 ($\sigma = 1.06$)
seldom	3.64 ($\sigma = 1.59$)

Table 4.1 indicates that regular users of VR-systems are more accurately pointing than those participants with less VR-experiences. However, the results are not statistically significant.

For each participant the mean pointing accuracy in the pointing task was calculated. These values served as a control measure to decide if the pointing person did not point at the correct object or if the observer did not see the correct object. In other words, the mean pointing accuracy helped to classify how accurate the pointing person really was.

In addition, the mismatches were counted that the observing subjects made when trying to figure out where the pointing person was pointing at for both conditions, real and virtual, over the three pointing object classes. The results are shown in Table 4.2.

Table 4.2: Mismatches in percent over the three pointing object classes for virtual and real dashboard.

object class	**real dashboard**	**virtual dashboard**
touch	0.0 %	30.0 %
outline	1.6 %	11.6 %
distance	13.3 %	25.0 %

As expected, it was very simple for the observer to decide where the pointing person is pointing at when directly touching the object on the real dashboard with his real finger (0.0% of mismatches). It was also simple for the observers to identify real objects for which the outline was drawn in the air (1.6% of mismatches). For the objects pointed at from a distance, the subjects had 13.3% of mismatches. However, one has to consider that no verbal communication was allowed.

For the virtual condition, the observers generated 30.0% of mismatches for directly touched objects. This is not surprising since it could be expected that it would be hard for the pointing participants to judge the position of their real finger in relationship to the virtual objects due to focus and convergence mismatch and occlusion effects. It is difficult for the pointing users to estimate the distance of their finger to the object since the finger occludes the object that should be touched. This effect can be seen by the values measured for pointing accuracy.

It is also described by Drascic and Milgram [1996] that if people see a displayed object in context with their own body, their perception changes. If the real hand and virtual object are in close proximity, the disparity depth cue will always deliver the information that the object is in front of the hand. At the same time, the accommodation depth cue tells the user that hand and object are at different depths. Drascic and Milgram [1996] report on perceptual issues that should be taken into account when working with stereoscopic displays. They describe several depth cues that influence stereoscopic viewing and perception. The influence of false or uncontrolled depth cues can be compensated by other consistent depth cues in the VR-system. The human visual system is able to adapt very quickly to mis-calibrated systems. However, if the user is forced to switch between real and virtual objects (e.g. when pointing with the real hand at virtual objects), this may lead to disorientation and misperception.

The subjects looked at the dashboard a little from above and from the side. Figure 4.5 shows the coordinate system in which the dashboard was positioned. The measured positions the subjects pointed at differed mostly in the x- and z-direction with respect to the virtual target object's center. About 57 percent of the subjects penetrated the target objects in x-direction. The rest did not touch it; their fingers were in front of the object. About 67 percent of the subjects pointed under the target objects. The rest pointed above it. These measurements were made relative to the object center, so it is possible that the objects were partly touched.

The measurements indicate that the observers' mismatches are mostly due to inaccurate pointing of the pointing users. When asked to directly touch an object, the adjustment process took a lot more time, thus an initial exact touch position could have possibly been discarded due to too much thinking and fine adjustment. Regarding outline and distant pointing, the pointing happened more intuitively and faster as in reality. It must be emphasized that all of the mismatches leading to the 30% error rate for directly touched objects had a deviation of maximal one centimeter from the object's center like the example in Figure 4.6 showing small buttons. In these cases, the mean pointing error was already larger than the object size. There were also mismatches that repeated between subjects. Surprisingly, ten of the subjects made no mistakes when guessing directly touched virtual objects. This is a significant fraction of the 24 participants and needs further investigation.

The outlined virtual objects were identified with 11.6% of mismatches and the objects, which are pointed at from a distance, with 25.0% of mismatches. These error rates indicate that it is easier to identify objects that are outlined from a distance than objects that are pointed at from a distance without drawing an outline, regardless of whether the pointing happens in real life or in virtual reality. The participants also had to rate how well they were able to estimate the position of their real index fingertip when pointing at virtual objects. They answered on a 1

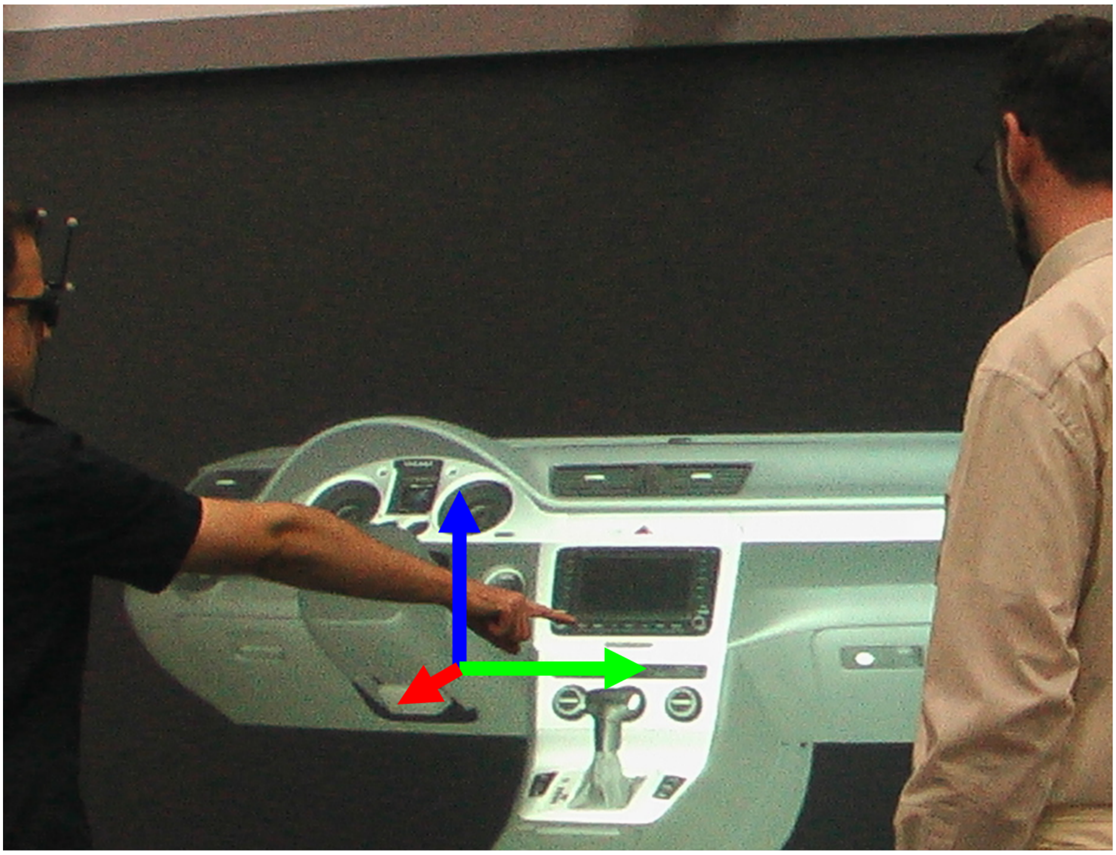

Figure 4.5: Alignment of the virtual dashboard in the coordinate system (x-axis in red, y-axis in green and z-axis in blue).

(very good) to 5 (very bad) Likert-scale and the mean answer was 2.65 ($\sigma = 0.98$), suggesting that this is not an easy task. Overall participants liked to point at virtual objects with their real hand: the mean answer was 2.25 ($\sigma = 0.85$).

Of course the use of virtual pointers, as introduced by Riege et al. [2006] with a pointing accuracy of nearly hundred percent would have been possible, but the intention was to evaluate pointing with the real hand, which is a very basic and intuitive interaction. Another advantage is that not all participants in a virtual scenario have to be equipped with pointing devices they possibly use for the first time. Especially for the outline and distance pointing, the effect of subjects guessing the target object simply by the coarse direction the finger pointed at could not be measured. To reduce this influence intentionally, objects were chosen that lay very close together as well as bigger objects with no close neighbors. It is furthermore expected that this effect occurs in reality as well as in virtuality.

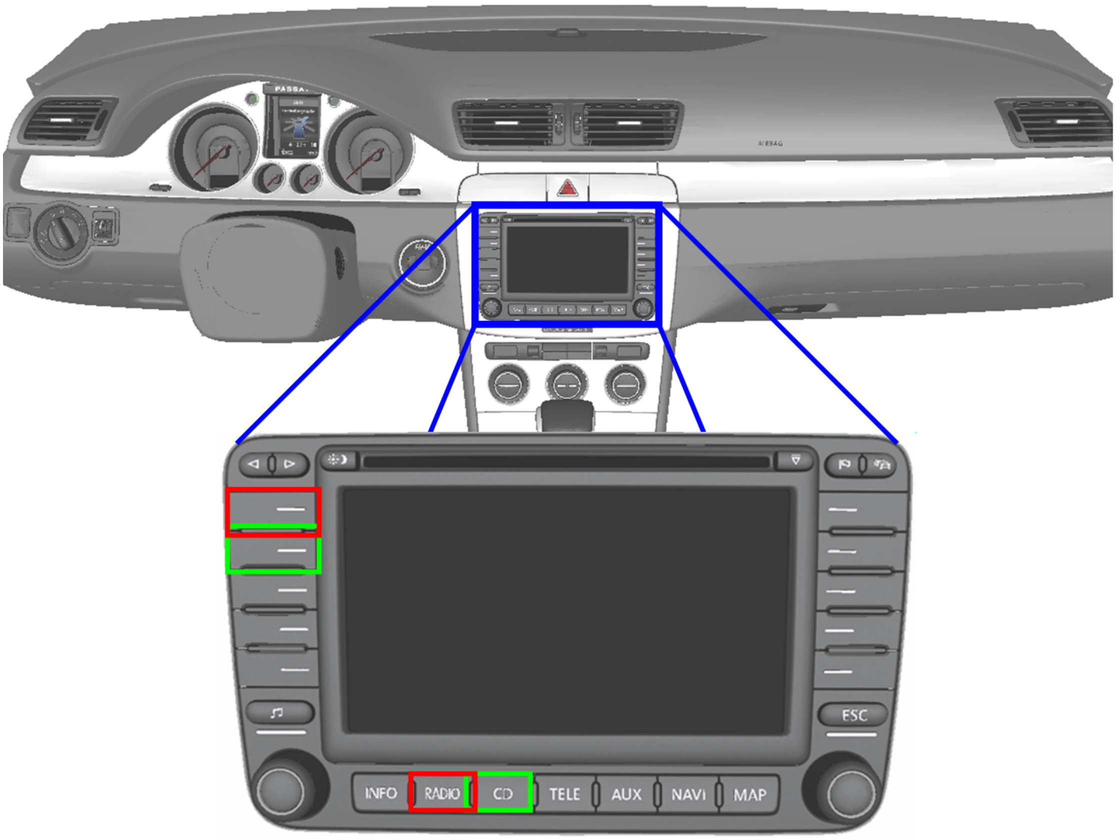

Figure 4.6: The buttons on the navigation system were the smallest objects in the pointing task. This area of the dashboard was displayed in the middle of the screen, which was also the best calibrated area of the four-projector system. The buttons with red frames are common mismatches for the neighboring buttons marked by green frames.

4.4 Conclusions

First results of an initial pointing experiment comparing real world pointing to pointing in a virtual environment were presented. In this experiment, participants were not allowed to verbally describe the object in addition to pointing at it, which is a somewhat artificial situation. The results indicate that outline pointing was the most secure way to help the observer identify the correct object.

Interactions that include directly touching virtual objects with the real hand need to be further investigated. In the experiments, these touch-based selections resulted in the largest number of mismatches, but there were also 40% of the users who could correctly identify the targets in this situation. Directly touched virtual objects smaller than two centimeters cannot be safely identified, which is trivial in the real world for most situations. The stereoscopic two-user system works quite well for common assembly scenarios involving larger objects. But, when it comes to interacting with a fully functional virtual navigation system equipped with small buttons, the use of a virtual pointer representing the user's finger is recommended.

Distance pointing is prone to error in reality as well, but it performs even worse in virtual scenarios. During informal tests with the two-user system, users pointed at objects and verbally described them as well. In these situations, no pointing inaccuracies were detectable. The famous "put-that-there" Bolt [1980] approach seems to work quite well even for smaller objects. However, this needs to be further investigated as well as the exact thresholds at which the different pointing techniques start to fail.

In some virtual environments, pointing often is triggered by symbolic or spoken commands and not by the arm of the user or an avatar representation. That is obviously not very natural and hard to understand by inexperienced users. Pointing gestures in real life happen intuitively and are almost always accompanied by verbal commands. That makes pointing gestures more precise and compensates uncertainties as shown by Wong and Gutwin [2010]. Due to this, the difference between virtual and real-world pointing should be even smaller when speech is not artificially excluded as in the pointing experiment in this thesis. Wong and Gutwin [2010] also conclude that it is essential that users do not have divergent views in the meaning of seeing different representations of the pointing gesture in relation to the objects in the environment. Therefore, users must be provided with individual stereoscopic views as in the two-user projection setup in this thesis.

According to Hindmarsh et al. [2000], a limited field of view in CVEs (e.g. a desktop or HMD) causes users to switch their views between referring gesture and referenced object since it is not possible to perceive both in one view. In the recent pointing experiment, the stereoscopic field of view provided by the stereo glasses was limited. But since the glasses did not cover the whole human field of view, users were still able to perceive the complete pointing gesture of the arm in their peripheral vision. This indicates that pointing gestures of the user's hands or arms can be better perceived in projection-based setups than in HMD-based setups.

Precise pointing is even more important when objects are referenced that are near or in a group of similar objects. This was also observed by Wong and Gutwin [2010] and Fraser et al. [1999]. Users often had problems identifying targets that were not obvious, like the buttons of the navigation system in the pointing experiment.

However, this is a general problem of pointing in both virtual and real worlds.

4.5 Summary

All presented pointing techniques are more error prone in virtual environments than they are in the real world. In particular, directly touched virtual objects smaller than two centimeters cannot be safely identified, which is trivial in the real world for most situations. Pointing with the real hand in the stereoscopic two-user system works quite well for common assembly scenarios involving larger objects. For smaller objects, it is recommended to use a virtual pointer representing the user's finger.

Chapter 5

Face-to-Face Assembly

At the end of the initial study in Chapter 3, the experts pointed out that it would be interesting to have two or more users looking from different sides at the car while they are collaboratively working. The projection-based system enables two users to look at a virtual model through the same window defined by the projection screen. This works well for scenarios in which both users are standing side-by-side in front of a virtual model such as the front or back of a car. Many collaborative assembly tasks require that two workers are partly facing each other to successfully perform the task. The initial study showed that it is hardly possible to simulate those tasks using projection-based setups even if they support multiple tracked users and multiple projection screens like a CAVE™. If the users would attempt to face each other, the projection of the virtual environment would always be occluded by their real bodies - except for cases wherein a table-top setup is sufficient. However, this is not the case for most automotive 1:1 scale simulations.

This chapter begins with a description of a two-user HMD-based setup simultaneously immersing two users in virtual face-to-face scenarios. This is followed by a study investigating assembly scenarios wherein two users are collaboratively working on the same object. Then the windshield assembly task is described in detail and the virtual as well as prop-based interaction techniques used for the evaluation of the windshield assembly task are introduced. It is investigated how stable the hand positions remain with the virtual method over the duration of the task and if the tangible prop enhances interaction and coordination by stabilizing the hands of the users. The evaluation section focuses on the user study with the questionnaire. Finally, the results are presented including the user performance and accuracy of the respective interaction method. Regarding the aspects of collaboration defined in the introduction of this thesis, this part takes into account individual stereoscopic views, communication with others by natural voice, perception of users' bodies by suitable avatar representations and collaborative interaction supported by passive haptic feedback and real world props.

5.1 HMD-Based Two-User Setup

For enabling face-to-face scenarios, a HMD-based two-user setup was implemented providing both a virtual surround view and an individual stereoscopic view. Since non-see-through HMDs were used, each user in the virtual world was represented by a basic avatar consisting of a virtual head and two virtual hands, exactly mapping their real hands' movements (Figure 5.1). Each user was able to see his own hands and the hands and head of the other user.

Figure 5.1: Each user was represented by a simple body model consisting of a head and two hands connected to the respective tracking target.

The two users' movements were restricted to a 2.80m x 2.50m optical tracking-frame with a 2.20m height. Two nVisorSX HMDs were used providing a resolution of 1280x1024 and an update rate of 60Hz, while covering a diagonal field of view of 60 degrees. For the hands, tracked ART-gloves with active LED-markers were used. The graphics update rate was 60Hz for all scenarios.

5.2 Study on Collaborative Assembly

This study investigates assembly scenarios wherein two users are collaboratively working on the same object. In reality, such a concurrent interaction of two people becomes necessary when relatively large and heavy objects need to be moved. Assembly tasks like this can be found in the production process of a car on the assembly line or more often during maintenance tasks in garages. Even during early development stages of a new car, it is necessary to perform simulations of the later production process considering it should be possible to efficiently assemble and disassemble a car. Such simulations are evaluated by assembly planners and ergonomists. Together they carefully design the assembly line processes following guidelines of ergonomics and effectiveness. Based on standard ergonomic simulations, which do not differ very much between different cars, a suitable real world scenario for collaborative assembly tasks was chosen. Such tasks are described in detail by process animations in CAD systems, which were used as guidelines to design the virtual scenarios. One assembly process usually consists of several steps, but for the task only the collaboratively performed process steps were extracted.

The windshield assembly is typically done by two workers using only one of their hands. They pick up the windshield from a rack next to the car by grasping it on each side using handles. The worker standing right grasps it with his left hand and the worker standing left uses his right hand. Then they carefully move the windshield over a distance of 2 meters to the front of the car and place it in its end-position. The intended task completion time (TCT) suggested by the CAD process animation is 20 seconds.

5.2.1 Interaction Techniques

Various approaches have been suggested to implement simultaneous manipulation of a single object by two or more virtual hands or users as by Cutler et al. [1997], Froehlich et al. [2000], Ruddle et al. [2002], Pinho et al. [2002] and Duval et al. [2006]. They implemented virtual interaction techniques without providing haptic feedback. In contrast to this previous work, the goal was to compare an unadulterated virtual method to a method based on a tangible two-user prop which provides a physical link between two users.

The prop-based method uses a simple aluminum pole with two attached handle

Figure 5.2: Snapshot from the video showing the process animation of the windshield assembly task.

bars to represent the windshield. They represent the object at least in one or two dimensions such as width, height and the balance point, but mainly they should exactly represent the grabbing positions the workers would use while manipulating the real object. The grabbing positions on the regarding virtual object were displayed as spheres. The windshield-prop weighs only about 3.5 kg, while the real windshield weighs about 14 kg, but at least the same centroid was provided. The pole prop also enables an additional communication channel as well as constraints between both users by transmitting forces via the pole.

The virtual method required that two users had to manipulate the windshield without the passive haptic feedback of the windshield prop via synchronized transformations using their tracked hands as inputs. Only visual feedback was provided by the movement of the virtual windshield. The windshield model was attached to the user's hand once the hand was in close proximity to the holding point represented by grab-spheres. The windshield was automatically released once the target position was reached. To achieve this behavior, passive nodes (p) were defined which are the grab-spheres on the object and active nodes (a) were defined which are the virtual hands. Valid grabs always consist of a link between an active and a passive node. For the detection of a grab, collision-spheres were used. If an active node is closer to a passive node than to a certain radius, (p-a $<$ r) a valid grab is established. The remaining distance between the nodes is stored and considered in the following calculations. If there is a defined minimum count of valid grabs, the object is grabbed. The object's position and orientation is calculated as follows:

- **Position:** Sum of positions of the active nodes causing the grab minus the relative start distance to the passive node divided by the number of nodes.
- **Orientation:** Result of a SLERP - operator between the orientation of the active nodes averaged by a factor of 0.5.

The grab is released if the object has reached its target position. Then all pairs of active and passive nodes are deleted.

Figure 5.3: Comparison of prop-based and virtual manipulation technique.

5.2.2 Evaluation

The stability of the hand positions over the duration of the task is an indicator of the reliability of the virtual evaluation of ergonomic aspects for the described assembly task. Thus, the question was how much the positions of the hands in the virtual condition varied with respect to the virtual windshield's holding points. The prop-based approach should result in quite fixed hand positions.

- H1: no stable hand positions with purely virtual interaction
- H2: the prop supports and enhances interaction by stabilizing the hands of the users

Visual dominance is the cause for an effect that could be observed in earlier VR-sessions: if a virtual object connected to the user's hand stops moving, the user also stops moving his hand even if there is no haptic feedback forcing him to stop. It was expected that the participants would try to keep their hands inside the holding spheres of the virtual handles on the object. That should lead to a stable distance between the user's hands holding the windshield over the duration of the task. It was speculated that visual dominance and verbal communication between users could compensate for missing haptic feedback.

Twenty subjects aged 25 to 45 (18 male, 2 female) volunteered to participate in the study. All of them had correct stereo vision. They had different backgrounds, including assembly planners, ergonomists and students with different virtual reality experience. A post questionnaire aimed at demographic issues, VR-experiences and the task itself was included. Items addressing collaboration by Biocca et al. [2001], involvement by Hofmann and Bubb [2003], awareness by Gerhard et al. [2001], co-presence by Schroeder et al. [2001], usability and preference were included as well. Some exemplary situations and questions:

- My partner worked with me to complete the task. *(collaboration)*
- I worked with my partner to complete the task? *(collaboration)*
- To what extent did events occurring outside the 3D scene distract from your experience in the virtual environment? *(involvement)*
- I enjoyed the virtual environment experience. *(involvement)*
- I was aware of the actions of my partner. *(awareness)*
- I was immediately aware of the existence of other participants. *(awareness)*
- How aware were you of the existence of your virtual representation? *(awareness)*
- When you continue to think back on the task, to what extent did you have a sense that you are together with your partner in the same room? *(co-presence)*
- Is it easy to use this technique for assemblies? *(usability)*
- I used the interaction technique successfully every time? *(usability)*

- It was easiest for me to coordinate my actions with my partner when I used *(preference)*
- I would recommend this interaction technique to a friend? *(preference)*
- This interaction technique is fun to use? *(preference)*

The questions were to be answered on a scale of 1 to 5 (1 = to a very small extent and 5 = to a very large extent). As a statistical measure, a one-way ANOVA was used to compare mean differences using a 95% confidence interval. At the end of the questionnaire space for general comments and remarks was provided.

The participants completed the task in pairs of two using the virtual method as well as the prop method. At the beginning, they were told to watch a video showing the task in order to get familiar with the movements they had to perform and how to manipulate the object (see Figure 5.2). They were instructed that it was important that they mimicked the task as precisely as possible. They were also able to complete one test trial in advance before the evaluation started. Each pair performed ten trials of each task and with each manipulation method. For each trial, the positions of the hands via the gloves and the center of the windshield during the manipulation as well as the time needed to complete the task were recorded.

After putting on the HMDs and the gloves, the participants were asked to stand in front of the car next to the rack with the windshield and to orient themselves in the virtual environment. They then picked up the windshield by its handles and began to move it to their target position. They had to walk about 2 meters while balancing the windshield by adjusting the positions and orientations of their hands (Figure 5.4). During the preparation of the scenario, a deviation of two degrees and one centimeter from the object center was identified as a suitable tolerance that still allows finding the target position in the virtual condition without frustration and with acceptable task completion times. The tolerance for the prop-based method was the same. As soon as the object was within the tolerance area near the target position, the object was colored dark-green. Then the participants had to hold this position for three seconds in order to prevent random matches, and the object's color changed to light-green and then froze into place. At this point, all measurements were stopped and the task was completed. Collisions between windshield and car were not considered.

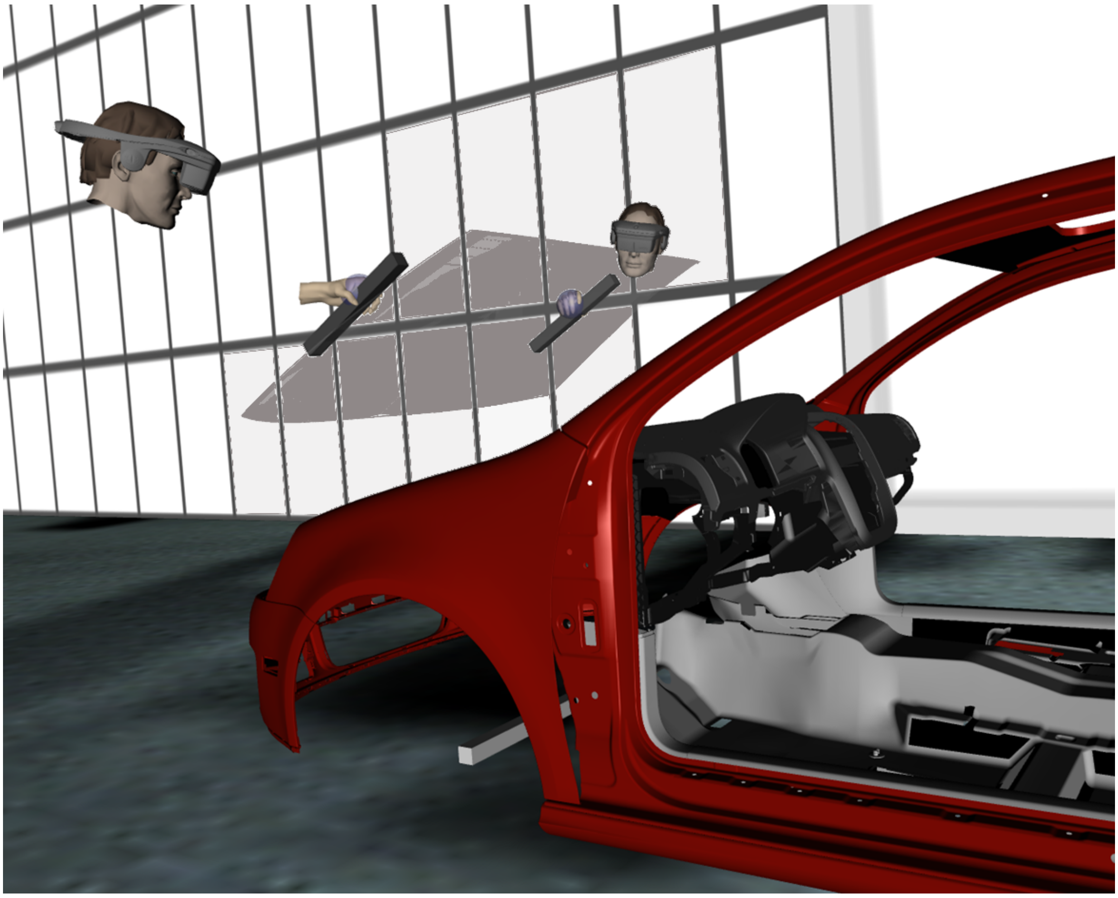

Figure 5.4: The basic avatars, the virtual windshield with its handles including small spheres representing the grab positions, the car and an assembly hall as background were composed in the virtual environment.

5.2.3 Results

Table 5.1 shows the presence measures for the collaborative HMD-scenario and the two different interaction methods. Collaboration and awareness were significantly better for the prop-based method, and involvement was still marginally better as well. Co-presence did not prove significant, which is reasonable considering that the visual representations of both users were identical for both tasks. At this point, it must be mentioned that the participants reacted with great pleasure when seeing each other through their HMDs; some of them waved at each other or even tried to shake hands.

As Table 5.2 indicates, the usability of the prop-based interaction was rated significantly higher than the usability of the purely virtual interaction. Most of the participants (93.75%) preferred the prop interaction over the virtual interaction method (6.25%).

Table 5.1: Presence measures for the assembly scenario corresponding to the used technique.

item	**prop**	**virtual**	**ANOVA**
collaboration	4.15 ($\sigma = 0.88$)	3.41 ($\sigma = 1.28$)	F(1,38)=4.05 p=0.009
involvement	4.12 ($\sigma = 0.80$)	3.60 ($\sigma = 1.23$)	F(1,38)=4.06 p=0.052
awareness	3.93 ($\sigma = 1.16$)	3.20 ($\sigma = 1.23$)	F(1,38)=3.98 p=0.010
co-presence	4.50 ($\sigma = 0.73$)	3.91 ($\sigma = 1.24$)	F(1,38)=4.30 p=0.166

Table 5.2: Mean differences regarding usability and preference for the two interaction techniques.

item	**prop**	**virtual**	**ANOVA**
usability	4.03 ($\sigma = 0.90$)	2.57 ($\sigma = 1.23$)	F(1,38)=3.91 p=0.001
preference	93.75 %	6.25 %	

The distance between the centers of the handles on the windshield prop was 93 cm. Table 5.3 shows that while using the prop-based method, the distance of the user's hands on the handles varied by 2.26 cm (1.13 cm per user), which may be due to slight hand adjustments without dropping the prop (Figure 5.5). Using the virtual method, this distance varied by 26.34 cm (13.17 cm per user). This obvious difference in measurements adds to hypothesis H1. But what's more surprising is that the average task completion time (TCT) was 17.16 sec for prop-based and 42.08 sec for the virtual condition, which is more than twice as much. A closer look at individual TCTs revealed that some of the participants were repeatedly able to reach times comparable to the prop manipulation. As previously mentioned, the intended TCT given by the CAD process animation is 20 seconds; so the prop method's TCT matches very well. Some participants were even faster because they started to get competitive and did not have to care about damages to the car. Altogether the only slightly differing distances between users' hands while using the prop-based method and the well-matched TCT add to hypthesis H2. The main

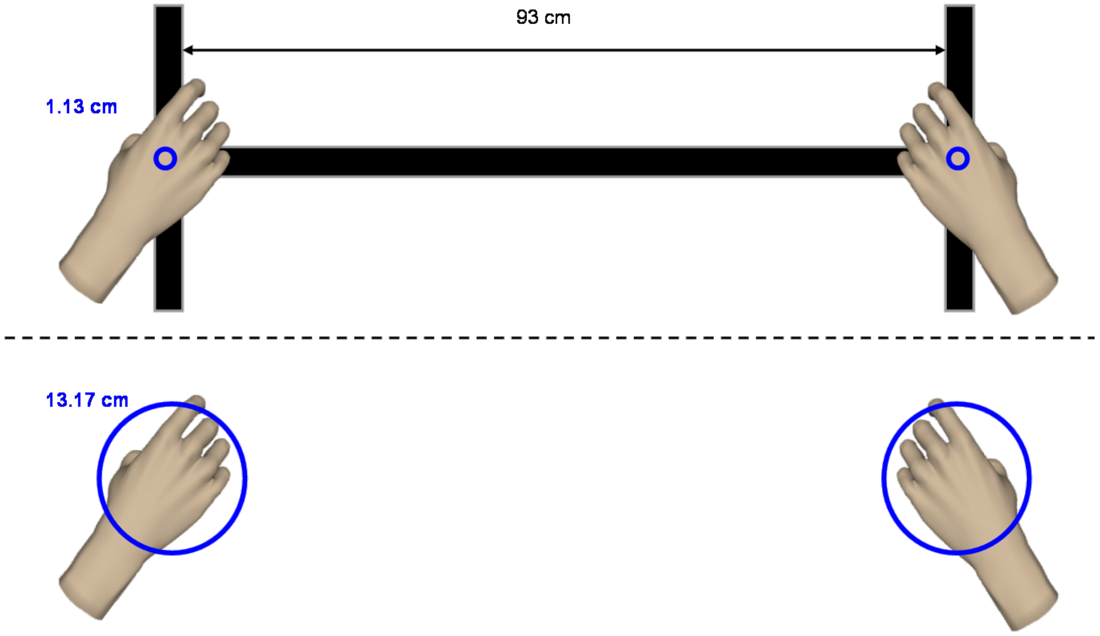

Figure 5.5: Average deviation of the hands from their initial position corresponding to the interaction method, visualized by blue spheres.

differences in TCT and hand positions were found in the fine adjustment phase close to the target position of the windshield. It is assumed that the participants needed a lot more time without the prop due to the missing force feedback, which could not be sufficiently compensated for by verbal communication and visual feedback. In general, the participants talked more to each other when using the virtual method to instruct their counterpart.

Table 5.3: ADDH is a measure for the average deviation from the default distance of users' hands for each technique and corresponding task completion times.

	prop	**virtual**
ADDH	2.26 cm ($\sigma = 0.86$)	26.34 cm ($\sigma = 5.70$)
avg TCT	17.16 sec ($\sigma = 6.14$)	42.08 sec ($\sigma = 24.86$)

The participants, all of whom knew each other, synchronized their movements by talking to one another. Mostly at the beginning, they counted "1, 2, 3" or similar commands to start the interaction. At the end during the fine adjustment of the windshield, which took the most time, they exchanged commands like "up, down, forward" to instruct their partner. They even realized if their partner was moving too fast between start and target position and would ask him to slow down. This type of communication adds to the findings of Otto and Roberts [2003] and Roberts et al. [2004], who observed that in non co-located collaborative tasks, verbal communication is a key feature. In contrast to Otto and Roberts [2003], users made natural use of verbal communication without being aware of physical communication devices due to the co-location. Even though the setup is co-located, there is not much of a difference when compared to non co-located setups due to the use of HMDs, with direct voice communication and the windshield prop being the exceptions.

5.2.4 Conclusions

The HMD-based assembly task was further discussed in a post evaluation. Pictures taken from the HMD-trials were shown to ergonomists and discussed to what extent they thought that ergonomic issues could be evaluated with such virtual methods. They believed that taking snapshots of users during the interaction could be very helpful to evaluate the body postures at certain time-steps. Even the 13 cm deviation of the hand position per user during the unadulterated virtual interaction with the windshield could be tolerable. However, the long task completion time of the purely virtual scenario is ergonomically critical and would distort the results of the ergonomic evaluations. The task duration is a particularly important factor for the ergonomics of a task. It is investigated how long workers have to rest in static poses carrying certain weights and how often they have to repeat the complete task over one working day. In the virtual scenario, no correct weights were simulated, but when looking at different postures, the ergonomists are able to judge how much weight a worker should carry and for how long he should remain in a particular pose.

The performance of the virtual interaction depends on the used algorithm to transfer user inputs in a synchronized way onto the virtual object. At the beginning when designing the virtual interaction, it was imaginable that a physical approach based on virtual springs connecting hands and object would be a good method to manipulate the windshield. But it soon became obvious that the method finally used, reacted to the user's forces more directly and felt more like interacting with a real windshield than a physical solution. Furthermore, a physical collision handling between car and windshield would have had the effect that a simple drop of the windshield close to the target position would have finished the task very quickly.

Usually a complete assembly process includes several steps, including side-by-side as well as face-to-face collaboration. So it happens that two workers are placing a large object into the desired position and then one of them has to fix it with a tool such as a screwdriver while the other one holds the part in position. While the basic usability of the projection-based setup is much better than for the HMD setup, only the HMD setup can be used for almost all task sequences. Nevertheless, a two-user CAVE™would already work for a much larger set of scenarios than the single projection screen used in this study, even though true face-to-face situations can be only simulated in a multi-user HMD system. This adds to the findings of Tan et al. [2001] who determined that the choice of display technology directly influences the interaction methods and interface design. This is extended by Mandryk et al. [2002] who presented seven technical factors of displays that influence collaboration in groups as referred to in the related work section of this thesis.

As already investigated by Hindmarsh et al. [2000], a limited field of view in CVEs, either desktop-based or HMD-based, is often compensated by an increased level of verbal communication between users. This effect could be observed as well during the solely virtual manipulation of the windshield when participants made use of spoken commands to coordinate their actions with their partner. On the other hand, verbal communication was reduced to a minimum when participants used the tangible prop to manipulate the windshield. So the limited field of view of the HMDs was compensated for by the forces and constraints enabled by the prop instead of by more talking. This enables users to exchange other information on the verbal channel, such as planning next steps of the task or small talk. Roberts et al. [2004] describe the effect that if work related verbal communication can be replaced by small talk, the feeling of co-presence is enhanced.

Pinho et al. [2002] conclude that cooperative techniques can provide increased performance and usability in difficult manipulation scenarios. However, single-user manipulation is simpler to use and understand for most manipulation tasks. They further state that the use of a cooperative technique is applicable to those situations in which cooperation allows the users to better control some DOFs that cannot be easily controlled with the single-user technique. When emulating real-world tasks (e.g. the windshield assembly task), in VR it is often inevitable to provide cooperative techniques beyond single-user interaction. Collaboration between two users is often more complex, thus making it all the more necessary for techniques to be even more comfortable and easy to understand. The implementation of tangible two-user props is an intuitive and precise method to emulate even complex interactions between two collaborating users.

5.3 Summary

Regarding assembly tasks, the findings of this chapter indicate that in this HMD-based face-to-face situation, untrained users rapidly became proficient with the task, and the prop-based method resulted in enough realism to validate ergonomic issues in a virtual environment. The haptic link and the constraints provided by the windshield prop where major factors in improving the coordination between two users. Due to missing force-feedback, the virtual method is much less suitable for ergonomic evaluations since the task duration times were much higher than in real life, and the two-user windshield mounting task could not be performed with enough precision.

Chapter 6

The Two-User Virtual Seating Buck

THE two-user virtual seating buck is a HMD-based system enabling discussions of novel car interface concepts. Although the two users are seated side-by-side, they are simultaneously able to discuss design issues of the car's interior face-to-face. This would not be possible in co-located projection-based systems since the counterpart's face disturbs and occludes the view on the virtual environment. Additionally, the limited field of view provided by the HMDs forces users to move their head a lot more (e.g. to focus on the hands and head of the other user to understand what he is looking at or referencing to). Another point is that investigations of the car interior require users to look around and to make use of the 360 degree view enabled by the HMD. In that way, face-to-face situations can happen unintentionally while focusing on the task. Nevertheless, the position and orientation of driver and co-driver is clearly side-by-side.

Similar to the face-to-face assembly scenario in the previous chapter, physical real-world items of the car interior are used to provide passive haptic feedback. In contrast to the single-user and two-user assembly props described before, the physical items here are more constrained with respect to their integration in the car body. They are simply supposed to give feedback to the user's fingers when touching switches or buttons.

The automotive industry is increasingly using virtual models instead of real prototypes as described by Bordegoni et al. [2007] and Monacelli et al. [2004]. While CAD systems and simulation packages are in regular use, the acceptance of virtual reality technology is rather limited so far. This is the case even though there seem to be many prime application areas for virtual reality technology, such as the study of car ergonomics and the testing of new human-machine interfaces in cars. One reason is the missing support for multiple users, since the test and evaluation of new interface concepts is typically done with two or more experts discussing various issues in front of or inside a real car or car mockup.

Based on an already existing single-user seating buck, the two-user virtual seating buck was developed (Figure 6.1, Figure 6.2), which allows two people to investigate new interface concepts in a virtual car interior. As described in the introduction of this thesis, this idea was born out of the fact that only one user was stereoscopically immersed in the virtual car interior while other people surrounding him were excluded. Now two users wear HMDs and take on the role of the driver and co-driver respectively. The seating buck is a minimal car mockup consisting of two seats, steering wheel, pedals and a few appropriately positioned interface elements such as the light switch, the air condition controls or the navigation device. These interface elements provide passive haptic feedback, while the virtual environment allows the user to explore (e.g. a new navigation device interface). Since two co-located users are present inside the virtual car, it becomes important to represent each user with an avatar such that they can see each others' actions. In an evaluation, it should be determined which fidelity is required for the avatars' visual appearance and the motion of the heads and hands to be acceptable for studying new car interfaces.

Figure 6.1: The two-user seating buck scenario. The image on the left side shows the hardware setup. The right side shows the same perspective for the virtual car model including the two avatars.

The main contributions of this work are the design and setup of a two-user seating buck system as well as the evaluation of potential avatar representations for collaborative evaluations of car interface prototypes. The seating buck provides the driver and co-driver with perspectively correct views through tracked HMDs. Both users' hands and fingers are tracked as well to support complex interactions with the car interior. Different head and hand models for representing the two users in this co-located scenario should be evaluated.

In this chapter, another HMD-based setup, the two-user virtual seating buck is presented. Beginning with an introduction of the setup and components, the main focus is on an evaluation of body representations since they are essential to visualize users in a setup based on non-see-through HMDs. Comparing two body models, one basic and a more sophisticated one, it should be answered how basic such a representation can be to be acceptable for automotive engineers such that interactions are perceived as being realistic. After introducing the body models, the test scenario - a car interior - is described as well as the actions users had to perform to evaluate the body models. Finally, a number of interesting observations are presented regarding the influence of embodiments in collaborative scenarios. The aspects of collaboration addressed in this part are individual stereoscopic views, communication with others by natural voice, perception of users' bodies by suitable avatar representations and collaborative interaction supported by passive haptic feedback.

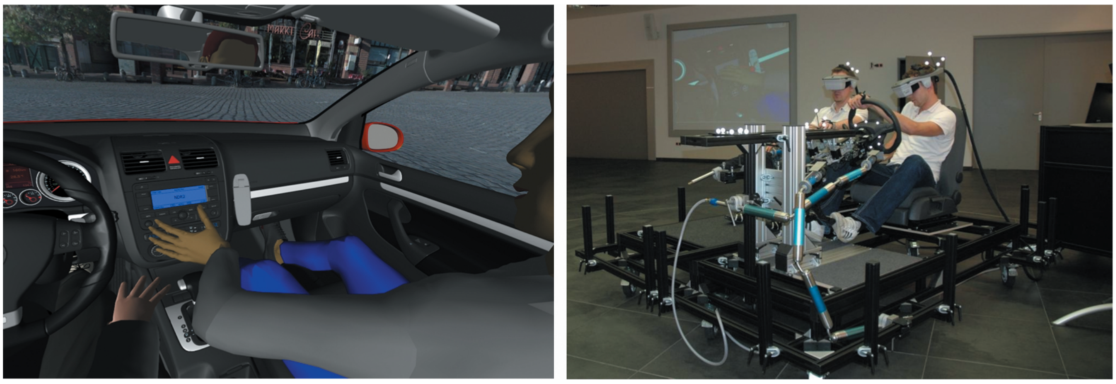

Figure 6.2: The right picture shows driver and co-driver in the real environment wearing their HMDs and discussing about functions of the navigation device. The left picture shows the corresponding virtual view of the driver.

6.1 Setup

The two-user virtual seating buck was set up as shown in Figure 6.2. The system consists of a minimal car-mockup, two HMDs for driver and co-driver, an optical tracking system and precise finger-tracking gloves. The seating buck is based on a chassis with two car seats, a steering wheel and real pedals. All of those physical items can be easily replaced by parts from other car types. The mockup can be adjusted in many ways (e.g. position and height of seats, steering column position and foot rest can all be modified). This makes it possible to represent a variety of cars, ranging in size from small to large.

The seating buck is framed by an aluminum rack where pneumatic mounts called "flexi-holders" can be attached. Each mount consists of two joints, which can be arranged and fixed in almost arbitrary configurations. The flexi-holders are used to hold and position prototype or real car parts such as navigation devices, touchscreen displays and several switches at their appropriate locations (Figure 6.3). These parts are typically tracked to insert them at the correct position into the virtual environment. The position of these passive haptic elements can be interactively changed during an evaluation session to adapt the setup to a new configuration of interface elements. Lok et al. [2003] state that providing passive or even active haptic feedback during the interactions increases the realism of the user's experience. In this case, real parts of a car (e.g. a light switch or navigation device prototype), are mounted at the appropriate locations to provide haptic feedback.

The tracking system consists of six cameras on the ceiling, which track the following objects:

- seating-buck, to align the virtual car model with the physical mockup
- steering wheel
- up to three moveable objects mounted on flexi-holders
- four finger-tracking gloves
- two HMDs

The virtual seating buck requires precise calibration since it is quite important that there is no penetration of virtual fingers and the virtual representation of a real car part while the user is touching the real part of the car.

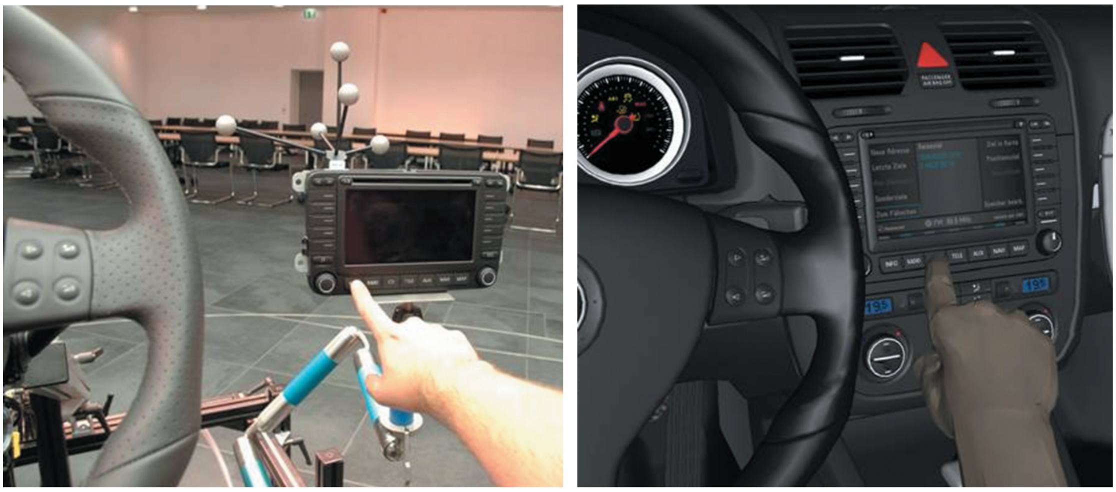

Figure 6.3: On the left side the navigation device panel mounted on an arm of the flexi-holder is shown and on the right side the corresponding virtual image through the driver's HMD.

The system is driven by an in-house software running two synchronized instances on separate computers. The functionality and interface of future car displays such as navigation devices and other multi-functional displays is typically prototyped in external programs running on separate PCs. The images generated by these display simulation programs are mapped onto polygons in the virtual environment. Events from real buttons and from fingers touching the display surface in the virtual environment are sent back to the simulations. This is done with early and inexpensive hardware models, which only have the buttons connected to a USB-controller. If a button is pressed, the controller sends key-events to the external display simulation (e.g. real fingertip and real button). Since both users are able to touch the same hardware prototype mounted on the flexi-holder (e.g. a navigation device), one user can see which virtual/real button the other is touching and how it changes the virtual display content. Alternatively, key-events can be generated by collisions between object bounding-boxes representing the virtual fingertip and virtual buttons. These events are sent from the VR system to the simulation PC. Thus fully functional car displays can be integrated into the virtual environment and operated as in a real car, including passive haptic feedback. Figure 6.4 shows the most important software and hardware components of the system in a schematic view.

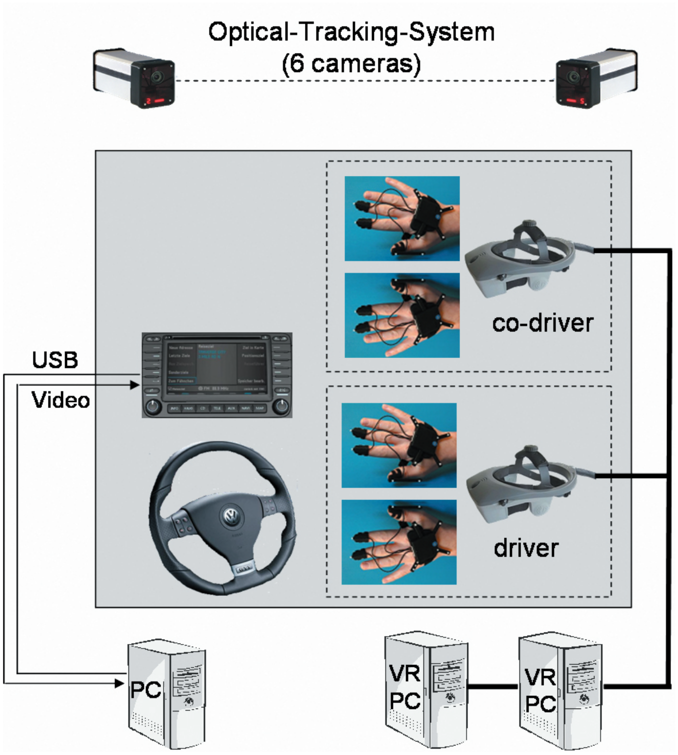

Figure 6.4: Schematic diagram of the two-user setup. Two separate PCs drive the HMDs. There is an additional PC for the simulation of the navigation device interface. Both hands of driver and co-driver as well as the HMDs and various other objects are tracked (steering wheel, objects on flexi-holders, etc.)

6.2 Study on Body Representations

The setup of the two-user seating buck allows two users to act as driver and co-driver while seated in the mockup and viewing a virtual car model. Natural voice communication is also possible and the shared virtual environment allows interactions by each user which should also be seen by the other. Since non-see-through HMDs have to be used, it becomes necessary to find a suitable virtual representation for each user.

The main question was how detailed the avatar representations need to be in the context of a virtual seating buck. Most important are the positions and orientations of the virtual hands and heads since they provide information for the actual focus of interest as well as the current action of each user. Therefore, it was necessary to track both hands and the head for each user in addition to the tracked car parts. Based on the tracking information, body models of different quality could be animated with lower or higher fidelity. The first study uses a simple body model and uses the tracking data to manipulate head and hands or arms with rigid body transformations. The second study compares the best result from the first study to a kinematic body model. For both studies, finger-tracking was not used since there was no finger-tracking hardware available for both users at the time the study took place.

6.2.1 Simplified Body Model

When developing new human machine interfaces in the automotive industry, the focus is on the effective and intuitive manipulation of car interface elements. Thus a basic visualization of head and hands and their movements might be sufficient for the evaluations of the new interface concepts. The question is how basic such a representation can be to be acceptable for automotive engineers such that the manipulations are perceived as being realistic.

Description of Model

The position and orientation of a simple head model or even only a model of the HMD worn by the user might be enough to show the current focus of interest and might be considered adequate. Since the head model did not support different facial expressions, it might be more appropriate to use a head model wearing an HMD, which hides a large part of the face and thus facial expressions could not be seen anyway. Therefore, three different head models were tested: simply an HMD, a male human head and a male human head wearing an HMD (Figure 6.5). Different hand/arm configurations were tested as well. In detail, just the hand, the forearm and the whole arm were implemented (Figure 6.6).

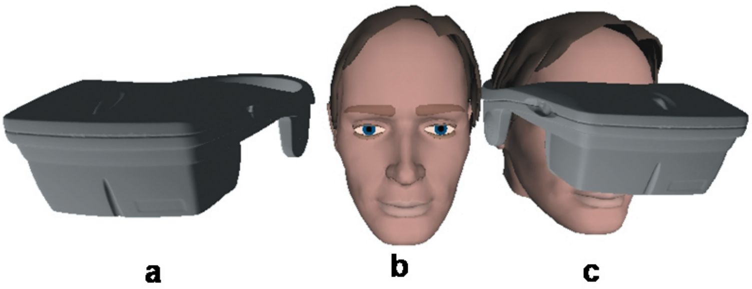

Figure 6.5: Head models for the first part of the user study. a) HMD model b) male human head c) male human head with HMD.

Figure 6.6: Arm/hand models for the first part of the user study. a) hand b) forearm c) arm.

Sanchez-Vives and Slater [2004] observed that it is disturbing for most users if they look down on their body and do not see it. Due to this, a male person sitting in the driver and co-driver seat was always displayed, each wearing a different outfit. These models were taken from the software Poser [2007]. The different head and hand/arm models were cut off from the static model in the driver and co-driver seat. These disconnected head and hand/arm models were connected to the head and hand tracking data and they therefore moved independently of each other and the rest of the body. Thus gaps between the different body parts can occur, which might destroy the illusion of a virtual human (see Figure 6.7). However, the focus was on the actions of the hands or the viewing direction of the head.

Figure 6.7: When the user leans forward and stretches his arm in order to reach the navigation device, the simplified body model produces gaps between the moving body parts and the fixed torso.

A car model was chosen, which was known to the participants of the study to avoid extreme distractions by focusing on the features of a new car model. The environment around the car was a static city model to make the participants feel like they were in a real car. A driving simulation would further enhance the realism, but the focus of this work was on the initial evaluation of novel interface concepts for the navigation devices and other multi-functional displays, which could be very well done in a stationary car.

Description of Tests

The study was designed to evaluate user preferences for the introduced representations of heads, hands and arms. Three different arm/hand models and three head models had to be compared. Thirteen users volunteered to participate in the study. Ten users had some experience with HMDs and VR, but they do not use such technology on a regular basis. The remaining three participants did not have any VR experience at all. All of them knew about the single-user seating buck or had even used it. The participants took place in the driver seat while wearing the HMD. The interviewer asked some questions about the participants' perception of their own avatar and they then had to evaluate the co-driver's avatar and actions while the co-driver showed the following interactions to each subject:

- Looking at the center console and touching buttons, including volume control and on/off-switch, of the navigation device using the left hand
- Looking up and touching the sunroof controls with the left hand
- Looking to the right and touching the power window switch of the co-driver's door with the right hand
- Looking into the eyes of the driver

For this study on body representations, a questionnaire was used and the participants were interviewed while experiencing the virtual environment to assess subjective ratings. As mentioned by Slater [2004], there are a lot more human factors that should be taken into account and questionnaires are too universal of a method to make reliable statements about a user's presence sensation in a virtual environment. The goal of this study was the evaluation of different body representations for the particular scenario instead of directly evaluating presence.

At the beginning of each interview, each participant was briefly shown the different arm/hand-models and the user had to rank the models with respect to their suitability for interface tasks. Then it was asked how much the participant likes the appearance of the model, followed by a questions regarding how good the movement of the model is. Both questions were rated on a one to five scale (1=very bad, 5=very good). Questions were asked for each arm. The ranking results, reflecting the first impression the participants had of the models, should establish a reference to approve the results from the scaling questions or to find differences.

The same procedure was carried out for the different head representations of the co-driver. These subjective ratings and rankings of the body parts were analyzed using a t-test for repeated measures to check the equality of means under a 95% confidence interval. It was also asked in a closed question ("yes" or "no") if the legs and feet should be moving. At the end of each interview, shortly after dismounting the HMD, the participants were asked questions on disorientation and 3D-experiences and the possibility to provide general comments:

- How do you rate the appearance of the virtual hand/arm/head representation?
- How do you rate the movement of the hand/arm/head representation?
- Do you think that the legs and feet should be moving according to the movement of your real feet and legs?
- To what extent did you feel disoriented?
- How do you rate your own experiences with 3D applications?

Table 6.1: Mean ranks for the examined basic head models (Friedman-Test).

	co-driver's head
male human head with HMD	1.54 ($\sigma = 0.52$)
male human head	1.85 ($\sigma = 0.90$)
HMD	2.62 ($\sigma = 0.65$)

Table 6.2: Mean ranks for the examined simple arm models (Friedman-Test).

	driver's arms	**co-driver's arms**
arm	1.46 ($\sigma = 0.66$)	1.62 ($\sigma = 0.77$)
forearm	2.15 ($\sigma = 0.80$)	2.00 ($\sigma = 0.82$)
hand	2.38 ($\sigma = 0.77$)	2.38 ($\sigma = 0.77$)

Results

Table 6.1 shows that the male human head with HMD was ranked best. So the option closest to reality, corresponding to the real co-driver wearing the HMD, was evaluated best and the minimal head representation showing only the HMD was the worst. Table 6.2 shows that the participants of the study favored the driver's full arm model over the forearm model. The hand model was ranked lowest. For the evaluation of the co-driver's arms, the ranking led to the same results with slightly different values.

Figure 6.8 shows the rated appearance and movement of the head models. The movement of the HMD compared to the male human head (p=.001) and compared to the male human head with HMD (p=.002) was rated significantly better. This is in contrast to the ranking of the different head models, which was provided after a brief look at the head models. In regards to the appearances, no significant differences were found.

Figure 6.9 shows that the appearance of the hand was significantly worse compared to the full arm model (p=.01). It is also indicated that the full arm model had the best movement and appearance.

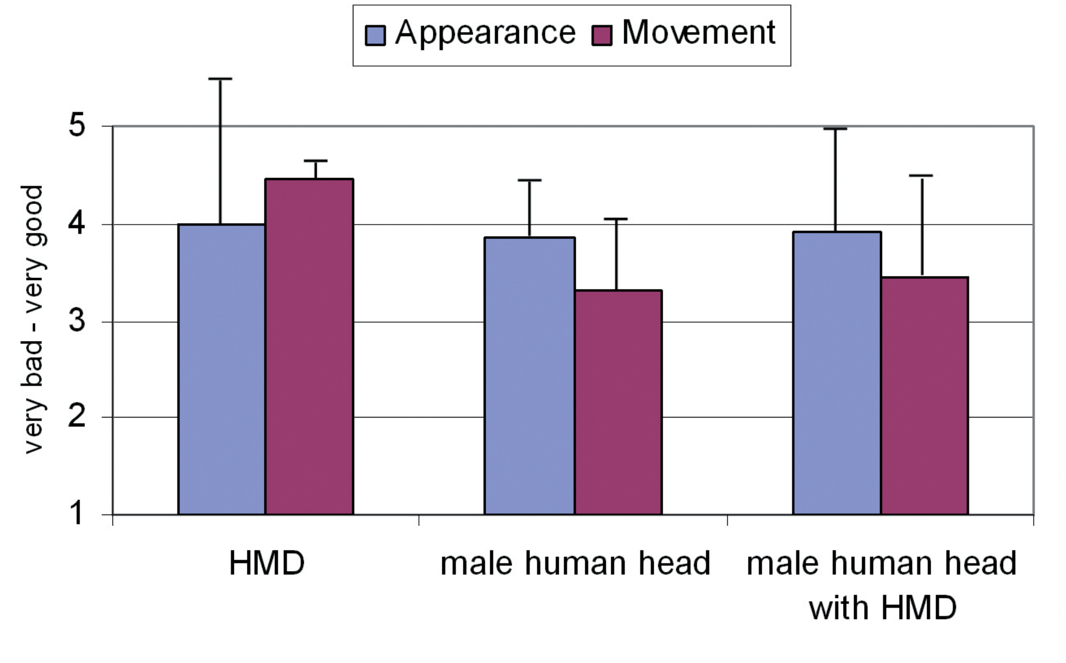

Figure 6.8: Subjective ratings of appearance and movement for the simple head representations.

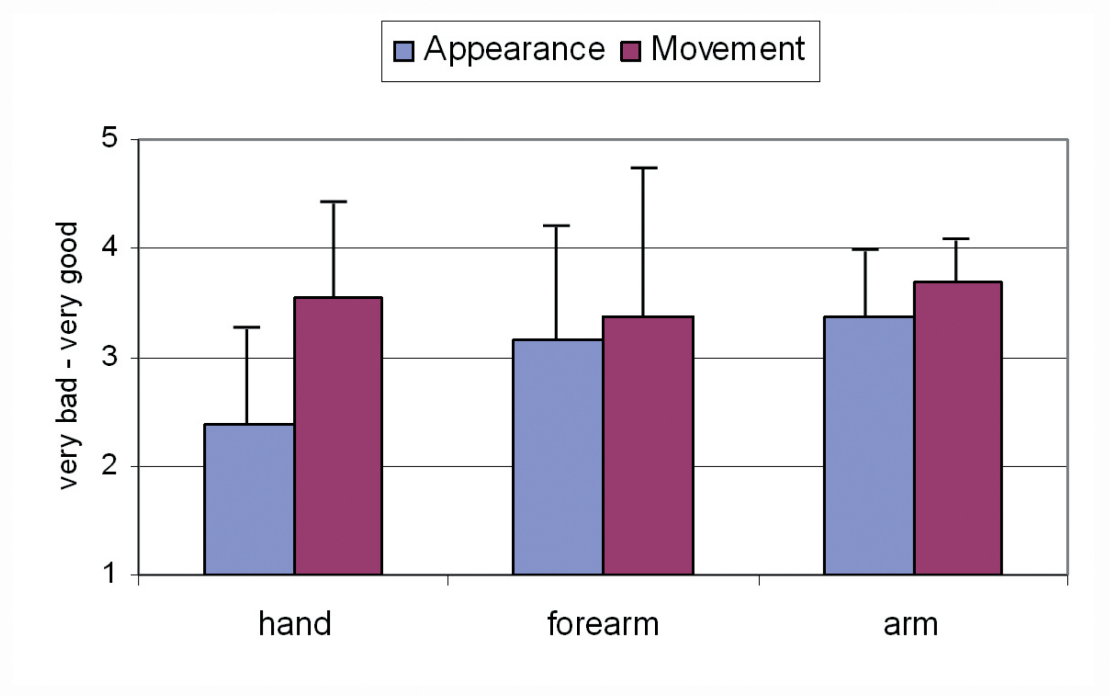

Figure 6.9: Subjective ratings of appearance and movement for the simple arm representations.

6.2.2 Kinematic Body Model

The first study mostly indicates that the more complete and naturally behaving the body model is, the more it is accepted for car interface studies - except for the head model, which will be discussed later. In the next step, a human body model was used which is controlled by inverse kinematics: the RAMSIS.

Figure 6.10: The RAMSIS model positioned in the co-driver seat.

Description of Model

The Ramsis [2007] is a complete human body model developed by Human Solutions, which is widely used for studies on car ergonomics and also during modeling of car interiors. The RAMSIS model can be parameterized to represent different body models, including the 1.90 meters tall and 95 kilograms weighing man (95 percent male) or a petite 57 kilograms weighing and 1.51 meters tall woman (five percent female). The interior of the cars is optimized in terms of ergonomics and visibility for occupants between these two extremes. This is the main application domain for the RAMSIS model, which is mostly used in CAD applications and real-time VR-systems.

The RAMSIS model was created from a body scan of the co-driver, who performed typical interaction tasks while sitting in the co-driver seat. Only one person was scanned because it would have been an unreasonably high effort to scan each person participating in the study. This meant that it was only possible to evaluate the RAMSIS-co-driver from the driver's perspective. The participant of the study, who was taking the driver's seat, was still represented by a simple Poser model. The RAMSIS was positioned in the co-driver seat and the model was fixed from foot to hip. The data from the three tracking targets for the two hands and the head was connected to the respective body parts of the RAMSIS model. The realistic movement of the rest of the upper body was computed based on the hand and head movements using kinematic laws. The RAMSIS simulation ran on a separate PC, which sends and receives information from the VR-system.

Description of Tests

The second user study was carried out among the same thirteen participants as the first one, with a few days in between. The participants assumed the role of the driver again, but this time their task was only to watch the co-driver acting. The representation of the co-driver alternated between the best ranked body model (full arm, male head with HMD, fixed torso and legs) of the first study and the presentation skin (most realistic looking) of the RAMSIS model. So this time only two pairs of arms and two heads had to be compared by ranking and by scaling questions. The virtual environment remained the same.

As in the first study at the beginning, both options were shown to get a first impression by ranking and to remember the model from the first study as well as to get to know the new RAMSIS model.

Table 6.3: Mean ranks for the comparison of the chosen simple body model with the RAMSIS (Friedman-Test).

	co-driver's arms
Ramsis arm	1.23 ($\sigma = 0.44$)
arm	1.77 ($\sigma = 0.44$)

	co-driver's head
Ramsis head	1.38 ($\sigma = 0.51$)
male human head with HMD	1.62 ($\sigma = 0.51$)

Results

The first impression of the RAMSIS model was rated better than the impression of the Poser model from the first study in regards to the arms as well as to the head (Table 6.3).

The scaling questions were asked after the participants had watched the co-driver acting. Figure 6.11 shows that there is no significant difference (p=.39) concerning the appearances of the arm models, but the movement of the RAMSIS arm was evaluated significantly better than that of the full arm (p=.003). Accordingly, the results for the head models revealed no significant differences for the appearance of the male human head with HMD and the RAMSIS head. The RAMSIS head movement was judged significantly better than the rigid body motion of the male human head with HMD (p=.007).

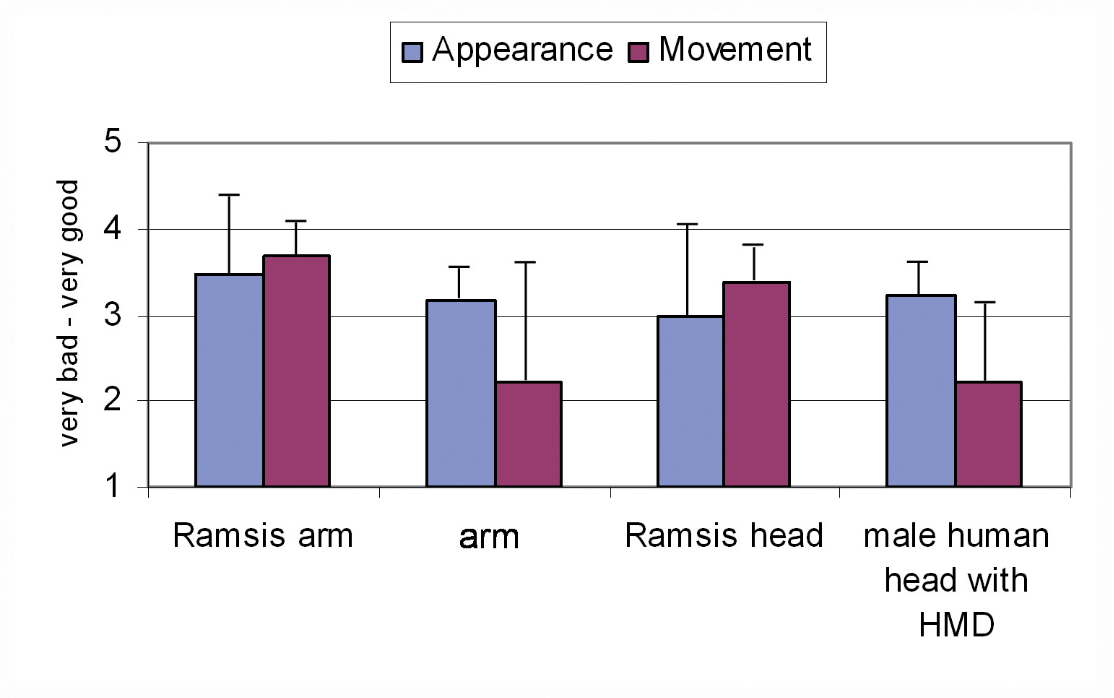

Figure 6.11: Subjective ratings of appearance and movement of body parts for the chosen simple body model compared to the RAMSIS.

6.2.3 Observations

One of the interesting results of the first study was the first rank for the male human head with HMD after the initial brief look at the co-driver avatar. The impression was that the participants still remembered the real person wearing the HMD and favored the avatar which looked most alike. In addition, facial expressions were mostly hidden and thus the correspondence with the facial expression of the real person did not play a major role.

Surprisingly, the plain HMD model was rated better for its movement than the other two head-models in the first study. It can be suspected that the head model's movement relative to a static body torso was considered less convincing than the abstract HMD model moving around independently. In addition, the HMD model did not provide any facial expression, which would not have matched the actual conversation between driver and co-driver.

In the second part of the study, the appearance of the male human head with HMD was rated better than the RAMSIS head, which might be evident of a too simplistic face model of the RAMSIS and the missing facial articulation. Participants who did not like the RAMSIS head also poorly rated the arm. This might indicate that the representations of both body parts are important in creating a convincing avatar.

The RAMSIS arm was clearly rated better on average with respect to appearance and in particular with respect to motion than the rigid arm model. This is a clear indication that the more realistic arm motion produced by the inverse kinematics model is considered important for evaluating the usability and ergonomics of various controls and interfaces in a car interior. It is also a clear statement that a simple six degree of freedom tracked hand or arm model is of limited use in such situations. One thing also observed in earlier tests was that it was disturbing for the users, if their right virtual arm was moving corresponding to their real arm, but the left arm was not. Thus it is necessary to track both arms even though for the driver (right-hand driving) the actions of the right arm are much more relevant than those of the left arm.

Concerning the inclusion of correctly moving legs and feet, the study delivered no reliable answer. Eight of the participants answered with "no" and the remaining five with "yes". This indicates that the majority did not consider articulated legs and feet to be of great importance even though the RAMSIS model would allow the simulation of these movements. However, such an avatar was not presented to the participants since it is quite difficult to receive a stable tracking signal in the foot space of the car due to occlusions.

All 13 participants had no problems with disorientation after they dismounted the HMD, and all of them rated their 3D-experiences as good. This is probably due to the high-resolution HMD (1280x1024), the precise tracking and calibration as well as the high update rates of 30 to 60 Hz. In addition, the virtual car was not moving during the studies, which reduces the susceptibility for disorientation and motion sickness. The participants did not complain about the HMD's limited field of view (60 degrees), which indicates that it does not play a major role for the evaluation of car interface concepts.

During the interviews the participants could also provide some general remarks about their experience with the system. One suggestion was to use a video camera to capture other participants in an evaluation session, who are outside of the car. The video stream should then be mapped onto an object, which can be seen by the driver and co-driver. The driver and co-driver could directly communicate with other participants, but they also wished to see them. It was also suggested to equip the other participants with HMDs as well. Some users wanted to get out of the car and move around the virtual car (e.g. to have a look at the engine bay). In principle, this is possible, but it would require to model the environment around the seating buck perfectly to avoid that users would run into obstacles. In addition, the cabling of the HMDs would have to be flexible enough to support such movements. Some subjects expressed that they would like to hear the sound of the environment to increase the realism. An almost obvious question was if it is possible to drive the virtual car. This would be very useful if someone wants to evaluate how navigation tasks, like entering an address while driving, can affect awareness to traffic. A driving simulation was also tested in the system but, as expected, quite a number of users suffered from motion sickness. It might be reasonable that the motion sickness is partially due to the limited field of view of only 60 degrees, but also due to the latency and update rate of the simulation system.

There were also some negative comments on the weight and limited ergonomics of the head-mounted displays. In particular, mounting the HMD while wearing the finger-tracking gloves was a difficult undertaking. The tracking system worked quite well overall. Sometimes the mapping of the real hand movements to the motion of the virtual hand was not perfect, which lead to collisions of finger markers with real objects in the car interior - in particular when users tried to interact like they are used to in a real car. This is a problem regarding hands of different sizes, since the used system could not be adjusted for different hand sizes. Users also wished to avoid the time consuming calibration procedure for the virtual hand model, which took about two to four minutes.

6.2.4 Conclusions

Overall, the two-user seating buck was well accepted and seen as a major improvement over the commonly used single-user seating buck. There were even some comments that passengers of the rear seats should be present as well or that some additional people should stand around the car and participate in the discussion.

The presented results add to the hypothesis of Garau et al. [2003] who presume that the more photorealistic an avatar is, the higher the demands for realistic behavior are. In this case, the simple Poser model can be seen as more realistic with respect to appearance than the RAMSIS, but it does not support realistic motions based on inverse kinematics. Overall, there is no big difference with respect to the rating of the appearance of both models, but the RAMSIS was clearly rated better with respect to movement. One may speculate that realistic behavior is more important than appearance and thus realistic motion compensates for less realistic appearance. Nevertheless, how well-articulated a human body model can be mostly depends on the available processing resources, so that in some tasks it is more helpful to go back to basic avatar representations if high task performance is required.

6.3 Summary

In this chapter, both the design and setup of a two-user seating buck system as well as the evaluation of potential avatar representations for collaborative evaluations in automotive scenarios was presented. In the context of the two-user seating buck, different head and hand models for representing the two users in this co-located scenario were evaluated. The findings indicate that the user representations and motions should be as realistic as possible even though the focus is on testing the interface elements operated by the users' fingers. The participants of the study also confirmed that the two-user seating buck is a major step towards the acceptance of virtual reality technology for interface studies in virtual car interiors, considering it allows the discussion and exploration between two people in a co-located virtual face-to-face situation.

Chapter 7

Conclusions

This thesis is concluded by a summarizing presentation of the main contributions related to the research questions defined in the introduction of this thesis. The aspects of multi-user systems are recapitulated with respect to their suitability to automotive tasks, as well as corresponding interaction techniques enabling collaboration in a group of users. Future work is discussed to round off the multi-user approach, indicating ideas for improvements and future research.

7.1 Contributions

Throughout this thesis, the evaluation of co-located collaborative interaction techniques for both projection-based and HMD-based stereoscopic two-viewer systems for automotive application scenarios was presented. The techniques were evaluated with the help of virtual emulations of real world scenarios involving two co-located collaborating users. The conducted studies overall confirmed that having a perspectively correct view of the virtual world is a basic requirement and premise for collaborative work in virtual environments. In general, co-located collaboration in automotive tasks involving face-to-face combined with side-by-side interactions is best supported by HMD-based multi-user setups augmented with multi-user props and passive haptic feedback. Due to better ergonomics, multi-user projection displays are well suited if the task requires only side-by-side interactions, and if gestures and postures of users' real bodies should still be perceived. Overall, the studies presented in this thesis confirmed that only stereoscopic multi-viewer systems enable natural interaction among co-located users for collaborative tasks within the reach of the users' arms.

To make collaborative meetings in co-located VR-setups successful, at least all participants of the group should be provided with basic VR-features. In this thesis, five aspects influencing collaboration in virtual environments were evaluated in the context of co-located automotive scenarios: vision, speech, referencing, embodiment and collaborative interaction. These five aspects directly correspond to the five main research questions defined in the introduction of this thesis, which are revisited below.

Which precision is required to enable natural gesturing and pointing? How is it possible to achieve the required precision? (referencing)

The possibility of pointing with the real hand at virtual objects enables the most natural and intuitive interaction metaphor, which is highly appreciated in projection-based collaborative virtual scenarios. The two-user stereo projection screen ensures that both participants of a VR-simulation converse about the same objects or problems at any time. It is essential that everyone can reference a specific object by simply pointing at it, which was not otherwise possible in a single-user system. Nevertheless, the findings show that very small objects that are directly touched cannot be identified without the chance of a mismatch. Scenarios involving objects larger than a few centimeters are a good choice for a projection-based multi-user system. Otherwise, more precise techniques such as virtual pointers or virtual fingers should be applied, taking into account the requirement of obtrusive tracking devices. These findings are a guideline for implementing multi-user interactions. However, further research is required to identify the factors that influence the accurate perception of such multi-user interactions in stereoscopic multi-viewer displays.

What are the appropriate multi-viewer tools and interaction techniques to enable true collaboration among automotive experts? (collaborative interaction)

In regards to the introduced collaborative interaction techniques, the prop-based methods are the best choice when implementing collaborative interaction in co-located virtual environments which require a high level of realism. Tracked single-user props offer the possibility to hand over a tool or to exchange tools with different functions, thus enabling natural interaction between users. They are able to switch their roles by simply exchanging input devices. Two-user props provide an interactive link between the two users and improve task performance and collaboration. Additionally, the subjective ratings confirmed that the participants preferred the prop-based over the purely virtual method. The passive haptic feedback and the direct connection between both participants make it an indispensable part of a variety of multi-user interaction tasks in virtual environments. Nevertheless, it is imaginable that with sufficient training, experienced actors are able to produce representative data for ergonomic evaluations even with the presented virtual manipulation method.

How does verbal communication influence collaboration in co-located multi-user setups? (speech)

As already investigated by Hindmarsh et al. [2000], a limited field of view in CVEs, either desktop-based or HMD-based, is often compensated by an increased level of verbal communication between users. This effect could also be observed during the solely virtual manipulation of the windshield when participants made use of spoken commands to coordinate their actions with their partner. On the other hand, verbal communication was reduced to a minimum when participants used the tangible prop to manipulate the windshield. In this way, the limited field of view of the HMDs was compensated for by the forces and constraints enabled by the prop instead of by more talking. This enables users to exchange other information on the verbal channel such as planning the next steps of the task.

Are avatar representations for the users required? If so, which avatar fidelity is required? (embodiment)

The design and implementation of a two-user seating buck scenario was presented, which enables the test and evaluation of new user interface concepts for cars in a virtual face-to-face scenario involving the driver and co-driver. The flexible setup allows the experimentation with different versions of real knobs, buttons, switches and display surfaces, while the virtual environment displays the appropriate actions and content of the connected interface simulations to both passengers. The main conclusion with respect to the required avatar representations is that the body movement should be based on inverse kinematics with at least tracked hands and head, as realized with the RAMSIS. This came as a slight surprise since it was expected that less sophisticated models might be sufficient. This assumption was based on the fact that the focus is on the interactions of the hands with the interface elements of the car. The participants of the study, who are actual experts involved in the car development process, clearly indicated that they would prefer sophisticated avatars. Even the appearance of the head should be as articulated as possible. Thus the two-user seating buck is a great application area for well articulated virtual humans.

Which scenarios are appropriate for which kind of multi-viewer technology? (vision)

The results presented in this thesis show that it should be carefully considered which type of display technology is suitable for a given task. Projection-based multi-user setups are mostly suited for tasks executed in a side-by-side fashion due to mutual occlusion of the displayed graphics. The only way to perfectly support face-to-face situations with projection-based systems is by using a truly distributed setup consisting of two separate projection systems as in the studies introduced by Heldal et al. [2005] or Roberts et al. [2004]. But in contrast to their work in a co-located situation, communication is not restricted by audio devices or network latencies which improve the ease of use of the system. HMD-based setups can be equally used for side-by-side and face-to-face situations, but in general they come with other shortcomings such as the discomfort of HMDs, limited field of view and an increased potential for motion sickness. Surprisingly in the HMD-scenarios, the participants performed very well and were moving relatively fast. Even their sensation of presence and co-presence was quite high, although the limited field of view should reduce perceptions in their peripheral vision. Wireless HMDs would be a great benefit to simulate more complicated assembly tasks wherein users have to cross over each other. HMD-based multi-user setups in combination with multi-user prop-based interaction turned out to be the most general and also a quite effective approach for co-located collaboration in automotive assembly scenarios.

Once users have realized the advantage of a two-user system over a single user system, whatever type of display system is chosen, they immediately suggested involving even more tracked users with individual perspectives into the virtual assembly scenario. Each user needs to be provided with an individual stereoscopic view as well as with a task-specific view. Thus further development of stereoscopic display technology supporting three or more users is important. However, these technical developments need to be accompanied by the corresponding development of effective and task-specific collaborative interaction techniques for multiple users as presented in this work.

7.2 Future Work

This thesis presented insights on the most important aspects of multi-user collaboration in co-located scenarios. The results presented are in some way a roadmap for future developments within this field of VR-research. The next passages include suggestions for follow-up studies based on what was presented in this thesis, as well as improvements or extensions to setups and techniques.

Side-by-Side with HMDs. The correlation between side-by-side collaboration in a projection-based setup and in an HMD-based setup was not explicitly investigated in this thesis. The assumption was that if the HMDs support face-to-face collaboration, they can be equally used for side-by-side scenarios since the users are only changing their position to each other and each HMD provides a 360 degree view by turning the heads. Indirectly, the virtual seating buck scenario also represents such a side-by-side situation based on HMDs. But even though the driver and co-driver are sitting side-by-side, they are simultaneously able to discuss design issues of the car's interior face-to-face. This would not be possible in co-located projection-based systems since the counterpart's face disturbs and occludes the view of the virtual environment. In the projection-based setup, the users are still able to visually perceive the real world surrounding them in their peripheral vision, because the stereo glasses cover only a limited part of the human field of view. In contrast, the HMDs clearly limit the field of view; therefore it might be worth investigating how this technical issue influences side-by-side collaboration.

Excursus on Distance Perception. Distance perception in virtual environments is a heavily discussed topic. Some studies already examined distance perception in virtual environments, which are comparably large. An example includes walking to different locations in one room or walking in virtual rooms with different sizes as introduced by Interrante et al. [2006] and Interrante et al. [2007]. Veridical distance perception in virtual environments is an important issue concerning realistic virtual simulations, especially those based on HMDs. In earlier tests with the virtual seating buck, several people mentioned that they have the impression that the dashboard is too far away on the co-driver side when looking from the driver's perspective. There is still not enough knowledge about factors that influence correct distance perception in the near field of smaller virtual environments, such as the interior of a car. Both the viewing parameters of the scene and the physical properties of the HMD have to match each other and must relate to the physiology of the respective user. Often the field of view in the scene does not match the field of view provided by the HMD. This could prove an advantage since the user is able to see more of the scene, but at the same time, the scene may appear distorted. The only way to convince users that what they see is correct in size is to provide a calibration procedure that allows the adjustment of the viewing parameters of HMD and scene to each individual user. The precondition for that is knowledge of all technical restrictions of the used HMD, provided by the manufacturer, and that the VR-software allows for the adjustment of the viewing parameters. Veridical perception of sizes is the most critical issue concerning decision reliability in automotive scenarios. To establish VR-technology in the long run for the automotive industry, correct size perception is essential, thereby serving as a premise for an increased VR-acceptance.

Haptics or Constraints. Regarding the windshield task performed in the two-user HMD-setup, the prop-based method, with faster task completion times and better ease of use, was clearly preferred over the purely virtual windshield manipulation. A question arising from this experiment was, if simply the presence of haptics or the presence of constraints makes a difference in the performance of the task. While using the pole-prop, the users were physically linked through enforced constraints, meaning that if one user pulls, the other must follow suit. In order to figure out if there is a difference, it could prove promising to set up an experiment comparing the already introduced method using the pole-prop to a method in which the users are just holding the handles in one of their hands without any link in between. Presumably, however, the pole-prop will still outperform this method. It might be more interesting to find out if there is a difference when comparing the virtual manipulation of the windshield, without any haptic feedback, to the method using only the handles that are not physically linked.

Re-configurable Props. For day-to-day use with automotive experts, it is recommended to provide a simple construction set for building props of commonly manipulated objects. This includes single-user props as well as multi-user props. The props do not have to look like the virtual objects they are linked with via the tracking system, but they should at least match the object's properties including grabbing points, size and centroid. A multi-user prop is not limited to a particular number of users concurrently manipulating it, since it is explicitly displayed to each user where the object can be grabbed. In that way, the object's movement is the result of a number of forces without loosing performance due to complicated algorithms or collision handling. Only position and orientation of the tracked object have to be updated in each user's rendering instance.

Props in Projection-based Setups. The prop-approach can be well applied in setups in which non-see-through HMDs are used since occlusions must not be considered. In projection-based setups, physical objects or bodies of users will occlude the virtual scene. To overcome this limitation and to enable the use of props even in setups like CAVEs™, props of transparent material might be a solution. It should be investigated if transparent props have an equally positive effect on collaboration or if distortions caused by the transparent material are too disturbing for users.

Collaboration Beyond Two Users. In the future, even more than two participants should be involved directly in the virtual seating buck scenario (Figure 7.1). Each user will be provided with an individual stereoscopic and task specific view. Today there remains the situation in which additional participants, besides the driver and co-driver, have to watch the session at a separate projection screen. Since these participants have typically different roles, the views should be tailored to the actual user. The ergonomist could see the RAMSIS model, while the designer sees the knits of the seats and realistic illumination. The electrical engineer may see the electrical circuits at the same time. Following the setup introduced in this work, this would mean equipping every additional participant with an HMD to provide individual stereoscopic views. Recently new display devices have become available that are mounted on the head like glasses but are non-see-through. They are mostly monoscopic, but at least they enable individual viewpoint changes for each participating user if they are tracked.

Another issue characteristic of co-located multi-user setups is investigated by Argelaguet et al. [2010]. In contrast to truly distributed multi-user setups, two or more users cannot take the same position in the scene at the same time (e.g. in front of a projection screen). This even includes a certain area of privacy around the user in which physical proximity might be perceived as uncomfortable. If a user wants to show a certain object to his colleagues, it might be occluded by other parts from their respective viewpoints. A natural human reaction would be to align with the view of the referencing user or his pointing arm, but this results in penetrations of the aforementioned area of privacy. Argelaguet et al. [2010] suggest a more elegant way to overcome this issue. They propose the use of show-through techniques to improve co-located collaboration of multiple users. In their approach, virtual objects being pointed at are shown through occluding objects with respect to the users' viewing directions. However, it needs to be investigated if such an artificial yet quite effective technique would be accepted by automotive experts.

Figure 7.1: Additional users participating in a virtual seating buck session provided with an individual stereoscopic view and visible to their colleagues via their avatars.

Involving multiple users with different expertises in the discussion process is essential in increasing the acceptance of virtual reality technology in automotive development meetings, thus making it a reliable decision platform in the future.

Equally important is the improvement in the ease of use of gloves, HMDs and stereo glasses.

7.3 Closing Remarks

This work provides scientific evidence that multi-viewer systems improve the usability of virtual reality applications for a number of automotive scenarios. The requirements for these scenarios as well as the limitation of multi-viewer systems are clearly identified. These findings contribute towards an increased acceptance of virtual reality technology for collaborative automotive development meetings. The basic properties of co-located multi-user systems, multiple individual stereoscopic views and communication by natural voice are effectively extended by supporting natural gestures, virtual embodiments and collaborative interaction techniques based on passive haptic feedback and props.

Automotive engineers and managers are not virtual reality experts. Providing correct stereo vision, unobtrusive hardware and intuitive collaborative interaction metaphors allow them to bridge the gap between the real and virtual worlds more easily. Passive haptic feedback provided either by physical items placed in a virtual environment or by single-user or multi-user props is clearly an important milestone. Multi-user props and passive haptic feedback may therefore be the key to successful simulation of collaborative tasks in virtual environments.

The results of this thesis may also be applicable to a variety of tasks in other fields of industry or research wherein people are willing or need to collaborate with the help of virtual reality. Particularly in industry, wherein everybody relies on commercial software packages, the findings will hopefully convince software vendors to integrate multi-user capabilities into their products in the near future.

Bibliography

Maneesh Agrawala, Andrew C. Beers, Bernd Froehlich, Pat Hanrahan, Ian McDowall, and Mark Bolas. The Two-User Responsive Workbench: Support for Collaboration Through Individual Views of a Shared Space. In *Proceedings of Computer Graphics 1997 (SIGGRAPH '97)*, pages 327–332, 1997.

F. Argelaguet, A. Kuhnert, A. Kulik, and B. Froehlich. Improving Collaboration in Co-located Interactions with Show-Through Techniques. In *IEEE Symposium on 3D User Interfaces*, pages 55–62, 2010.

Kevin Arthur, Timothy Preston, Russell M. Taylor, Frederick P. Brooks, Mary C. Whitton, and William V. Wright. Designing and Building the PIT: a Head-Tracked Stereo Workspace for Two Users. Technical report, University of North Carolina at Chapel Hill, 1998.

Adrian Bangerter and Daniel M. Oppenheimer. Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture*, 6(1):85–102, 2006.

Steve Benford, John Bowers, Lennart E. Fahlen, Chris Greenhalg, and Dave Snowdon. User Embodiment in Collaborative Virtual Environments. In *Proceedings of CHI'95*, volume 1, pages 242–249, 1995.

Frank Biocca, Chad Harms, and Jenn Gregg. The Networked Minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual International Workshop on Presence*, pages 9–11, 2001.

Richard A. Bolt. Put-That-There: Voice and gesture at the graphics interface. In *Computer Graphics and Interactive Techniques*, volume 14, pages 262–270, 1980.

Monica Bordegoni, Giandomenico Caruso, Francesco Ferrise, and Umberto Giraudo. A mixed environment for ergonomic tests: tuning of the stereo viewing parameters. In *Eurographics Italian Chapter Conference*, pages 127–134, 2007.

Tolga K. Capin, Igor S. Pandzic, Hansrudi Noser, Nadia M. Thalmann, and Daniel Thalmann. Virtual Human Representation and Communication in VLNET. In *IEEE Computer Graphics and Applications*, volume 17, pages 42–53, 1997.

F. Caputo, G. Di Gironimo, G. Monacelli, and F. Sessa. The design of a virtual environment for ergonomic studies. In *Proceedings of XII Convegno ADM Design Tools and Methods in Industrial Engineering*, Rimini, 2001.

Jeff Chastine, Kristine Nagel, Ying Zhu, and Mary Hudachek-Buswell. Studies on the Effectiveness of Virtual Pointers in Collaborative Augmented Reality. In *Proceedings of IEEE Symposium on 3D User Interfaces 2008*, pages 117–124, 2008.

Andrea Corradini and Philip R. Cohen. Multimodal Speech-Gesture Interface for Handfree Painting on a Virtual Paper Using Partial Recurrent Neural Networks as Gesture Recognizer. In *Proceedings of the International Joint Conference on Neural Networks*, pages 2293–2298, 2002.

Lawrence D. Cutler, Bernd Fröhlich, and Pat Hanrahan. Two-handed direct manipulation on the responsive workbench. In *SI3D '97: Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 107–114, New York, NY, USA, 1997. ACM. ISBN 0-89791-884-3.

Neil A. Dodgson. Variation and extrema of human interpupillary distance. In *Proceedings SPIE*, volume 5291 of *Stereoscopic Displays and Virtual Reality Systems*, pages 36–46, 2004.

David Drascic and Paul Milgram. Perceptual issues in augmented reality. In *Proceedings SPIE*, volume 2653 of *Stereoscopic Displays and Virtual Reality Systems*, pages 123–134, 1996.

Andrew T. Duchowski, Nathan Cournia, Brian Cumming, Daniel McCallum, Anand Gramopadhye, Joel Greenstein, Sajay Sadasivan, and Richard A. Tyrrell. Visual Deictic Reference in a Collaborative Virtual Environment. In *Symposium on Eye tracking research and applications*, 35-40, 2004.

Thierry Duval, Anatole Lecuyer, and Sebastien Thomas. SkeweR: a 3D Interaction Technique for 2-User Collaborative Manipulation of Objects in Virtual Environments. In *IEEE Symposium on 3D User Interface 2006*, pages 71–74, 2006.

M. Fraser, S. Benford, J. Hindmarsh, and C. Heath. Supporting awareness and interaction through collaborative virtual interfaces. In *UIST'99*, pages 27–36, 1999.

B. Froehlich, H. Tramberend, A. Beers, M. Agrawala, and D. Baraff. Physically-based Manipulation on the Responsive Workbench. In *Proceedings of IEEE Virtual Reality 2000*, pages 5–11, 2000.

Bernd Froehlich, Roland Blach, Oliver Stefani, Jan Hochstrate, Joerg Hoffmann, Karsten Klueger, and Matthias Bues. Implementing Multi-Viewer Stereo Displays. In *WSCG (Full Papers)*, pages 139–146, 2005.

M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogny, A. Steed, and M. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proc. CHI 2003*, pages 529–536. ACM Press, 2003.

Christian Geiger and Oliver Rattay. TubeMouse - A Two-Handed Input Device for Flexible Objects. In *3DUI '08: Proceedings of the 2008 IEEE Symposium on 3D User Interfaces*, pages 27–34. IEEE Computer Society, 2008.

M. Gerhard, D. Moore, and D. Hobbs. Continuous Presence in Collaborative Virtual Environments: Towards the Evaluation of a Hybrid Avatar-Agent Model for User Representation. In *International Conference on Intelligent Virtual Agents*, pages 137–155, 2001.

Ilona Heldal, Ralph Schroeder, Anthony Steed, Ann-Sofie Axelsson, Maria Spante, and Josef Wideström. Immersiveness and Symmetry in Copresent Scenarios. In *IEEE Virtual Reality 2005*, pages 171–178, 2005.

K. Hinckley, R. Pausch, J. H. Downs, D. Proffitt, and N. Kassell. The Props-Based Interface For Neurosurgical Visualization. In *MMVR5: Medicine Meets Virtual Reality 5*, pages 552–562, Amsterdam, Netherlands, 1997. IOS Press.

Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. Passive real-world interface props for neurosurgical visualization. In *Proceedings of ACM CHI'94 Conference on Human Factors in Computing Systems*, pages 452–458, 1994.

J. Hindmarsh, M. Fraser, C. Heath, S. Benford, and C. Greenhalgh. Object-Focused Interaction in Collaborative Virtual Environments. *ACM Transactions on Computer-Human Interaction*, 7:477–509, 2000.

Jan Hofmann and Heiner Bubb. Presence in industrial virtual environment applications - susceptibility and measurement reliability. *Being There: Concepts, Effects and Measurement of User Presence in Synthetic Environments*, pages 237–246, 2003.

Victoria Interrante, Lee Anderson, and Brian Ries. Distance Perception in Immersive Virtual Environments, Revisited. In *IEEE Virtual Reality 2006*, pages 3–10, 2006.

Victoria Interrante, Brian Ries, Jason Lindquist, and Lee Anderson. Elucidating the Factors that can Facilitate Veridical Spatial Perception in Immersive Virtual Environments. In *IEEE Virtual Reality 2007*, pages 11–18, 2007.

Pieter Jorissen, Maarten Wijnants, and Wim Lamotte. Dynamic Interactions in Physically Realistic Collaborative Virtual Environments. In *IEEE Transactions on visualization and Computer Graphics*, volume 11, pages 649–660, 2005.

Benjamin Lok, Samir Naik, Mary Whitton, and Frederick P. Brooks. Effects of Handling Real Objects and Avatar Fidelity on Cognitive Task Performance in Virtual Environments. In *IEEE Virtual Reality 2003*, pages 125–132, 2003.

Regan L. Mandryk, Stacey D. Scott, and Kori M. Inkpen. Display Factors Influencing Co-located Collaboration. In *Conference Supplement to Computer-Supported Co-operative Work*, 2002.

D. M. Margery, B. Arnaldi, and N. Plouzeau. A general framework for cooperative manipulation in virtual environments. In *Proceedings of EGVE 99: Eurographics Workshop on Virtual Environments*, pages 169–178. Springer, 1999.

Gennaro Monacelli, Fabrizio Sessa, and Alessandro Milite. An Integrated Approach to Evaluate Engineering Simulations and Ergonomic Aspects of a New Vehicle in a Virtual Environment: Physical and Virtual Correlation Methods. In *FISITA 2004 30th World Automotive Congress*, pages 23–27, Barcelona, 2004.

J. Nielsen. *Usability Inspection Methods.* John Wiley & Sons, NY, 1994.

Oliver Otto and David Roberts. Importance of Communication Influences on a Highly Collaborative Task. In *Proceedings of the 7th IEEE International Symposium on Distributed Simulation and Real Time Applications*, pages 195–201, 2003.

Marcio S. Pinho, Doug A. Bowman, and Carla M.D.S. Freitas. Cooperative Object Manipualtion in Immersive Virtual Environments: Framework and Techniques. In *VRST 2002*, pages 171–178. ACM, 2002.

Poser. http://www.e-frontier.com/go/products, July 2007.

Jennifer J. Preece, Yvonne Rogers, and Helen Sharp. *Interaction Design: beyond human computer interaction.* Wiley & Sons, 2002.

Jiten Rama and Judith Bishop. A survey and comparison of CSCW groupware applications. In *SAICSIT '06: Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, pages 198–205, 2006.

Ramsis. Human Solutions RAMSIS Automotive. http://www.human-solutions.com/automotive industry/ramsis automotive en.php, July 2007.

Kai Riege, Thorsten Holtkamper, Gerold Wesche, and Bernd Froehlich. The Bent Pick Ray: An Extended Pointing Technique for Multi-User Interaction. In *3DUI '06: Proceedings of the 3D User Interfaces*, pages 62–65. IEEE Computer Society, 2006.

David Roberts, Robin Wolff, Oliver Otto, Dieter Kranzlmueller, Christoph Anthes, and Anthony Steed. Supporting Social Human Communication between Distributed Walk-in Displays. In *Proceedings of VRST'04*, pages 81–88, 2004.

Tom Rodden. A Survey of CSCW Systems. *Interacting with Computers*, 3:310–353, 1992.

Roy A. Ruddle, Justin C.D. Savage, and Dylan M. Jones. Symmetric and Asymmetric Action Integration During Cooperative Object Manipulation in Virtual Environments. *ACM Trans. on Computer-Human Interaction*, 9(4):285–308, 2002.

Maria V. Sanchez-Vives and Mel Slater. From Presence Towards Consciousness. In *8th Annual Conference for the Scientific Study of Consciousness*, July 2004.

Dieter Schmalstieg, Anton Fuhrmann, Gerd Hesina, Zsolt Szalavari, L. Miguel Encarnacao, M. Gervautz, and W. Purgathofer. The Studierstube Augmented Reality Project. In *Presence Teleoperators and Virtual Environments*, volume 11, pages 33–54, 2002.

Ralph Schroeder, Anthony Steed, Ann-Sofie Axelsson, Ilona Heldal, Asa Abelin, Jodef Wideström, Alexander Nilsson, and Mel Slater. Collaborating in networked immersive spaces: as good as being there together? *Computers and Graphics*, 25:781–788, 2001.

Mel Slater. How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments. In *Presence Teleoperators and Virtual Environments*, volume 13, pages 484–493, 2004.

Mel Slater and Martin Usoh. Body Centred Interaction in Immersive Virtual Environments. In *Artificial Life and Virtual Reality*, pages 125–148. John Wiley and Sons, 1994.

Desney S. Tan, Ivan Poupyrev, Mark Billinghurst, Hirokazu Kato, Holger Regenbrecht, and Nobuji Tetsutani. The Best of Two Worlds: Merging Virtual and Real for Face-to-Face Collaboration. In *IEEE International Conference on Multimedia and Expo*, pages 861–864, 2001.

Daniel Thalmann. The Role of Virtual Humans in Virtual Environment Technology and Interfaces. In *Frontiers of Human Centered Computing*, pages 27–38. Springer London, 2001.

Nelson Wong and Carl Gutwin. Where Are You Pointing? The Accuracy of Deictic Pointing in CVEs. In *CHI*, pages 1029–1038, 2010.

Curriculum Vitae

Personal

Name : Holger Salzmann
Birth Date : June 18, 1979
Birth Place : Bad Langensalza/Thuringia, Germany

Education

1986 – 1991 : POS „A. Diesterweg“, Ebeleben
1991 – 1998 : Gymnasium „Geschwister Scholl“, Sondershausen
1999 – 2005 : Diploma study in Computer Science at Technische Universität Ilmenau, Fakultät für Informatik und Automatisierung

Professional

2005 – 2008 : Phd Student, Volkswagen AG
2008 – 2010 : Electronics Engineer, Carmeq GmbH, Member of Volkswagen Group

Publications

H. Salzmann and B. Fröhlich. The Two-User Seating Buck: Enabling Face-to-Face Discussions of Novel Car Interface Concepts. In *Proceedings IEEE Virtual Reality 2008 Conference*, pages 75–82. IEEE Computer Society, 2008.

H. Salzmann, M. Möhring and B. Fröhlich. Virtual vs. Real-World Pointing in Two-User Scenarios. In *Proceedings IEEE Virtual Reality 2009 Conference*, pages 127–130. IEEE Computer Society, 2009. (Best Short Paper Award)

H. Salzmann, J. Jacobs and B. Fröhlich. Collaborative Interaction in Co-Located Two-User Scenarios. In *Proceedings Joint Virtual Reality Conference 2009*, pages 85–92. Eurographics Association, 2009.