Natural Interaction Metaphors for Functional Validations of Virtual Car Models

Mathias Moehring, Bernd Froehlich

Abstract— Natural Interaction in virtual environments is a key requirement for the virtual validation of functional aspects in automotive product development processes. Natural Interaction is the metaphor people encounter in reality: the direct manipulation of objects by their hands. To enable this kind of Natural Interaction, we propose a pseudo-physical metaphor that is both plausible enough to provide realistic interaction and robust enough to meet the needs of industrial applications. Our analysis of the most common types of objects in typical automotive scenarios guided the development of a set of refined grasping heuristics to support robust finger-based interaction of multiple hands and users. The objects' behavior in reaction to the users' finger motions is based on pseudo-physical simulations, which also take various types of constrained objects into account. In dealing with real-world scenarios, we had to introduce the concept of Normal Proxies, which extend objects with appropriate normals for improved grasp detection and grasp stability. An expert review revealed that our interaction metaphors allow for an intuitive and reliable assessment of several functionalities of objects found in a car interior. Follow-up user studies showed that overall task performance and usability are similar for CAVE- and HMD-environments. For larger objects and more gross manipulation, using the CAVE without employing a virtual hand representation is preferred, but for more fine-grained manipulation and smaller objects the HMD turns out to be beneficial.

Index Terms— H.5.2 [User Interfaces], I.3.7 [3-D Graphics and Realism], B.4.2 [Input/Output Devices], H.3.4 [Systems and Software]

1 INTRODUCTION

THE development processes of today's automotive industry are very cost- and time-intensive. Virtual reality technology has the potential to make evaluations of computer-generated models possible in very early phases of the process, which could significantly reduce the number of required hardware mockups as well as the number of design iterations. In particular, the visual validation of virtual cars has become accepted for several purposes throughout the product life cycle. However, for validating functional aspects, engineers still prefer the usage of hardware mockups. Existing virtual reality applications for such evaluations, e.g. concerning the examination of ergonomics issues or assembly tasks, suffer from abstract interaction metaphors and unrealistic object behavior. Natural Interaction – the way people are used to interacting with objects in reality – is not yet completely supported by current virtual reality systems.

For some time we have tried to use various rigid body simulations to implement realistic hand-object interaction in car interiors. While this is certainly the desired solution in our context, it turned out that current physics engines do not provide enough robustness to properly handle our complex geometries and constraints. Instead we developed a set of pseudo-physical interaction metaphors, which enable finger-based interaction with various objects typically found in car interiors. Our approach is based on a thorough analysis of the predominant types of objects, their constraints and the typical grasps.

- *M.Moehring is with the Group Research of Volkswagen AG, Wolfsburg, Germany, D-38436. E-mail: mathias.moehring(at)volkswagen.de.*
- *B.Froehlich is with the Virtual Reality Systems Group of Bauhaus-Universität Weimar, Germany, D-99423, E-mail: bernd.froehlich(at)uni-weimar.de*

Our main contributions are:

- The development of extended grasping heuristics, which work for multiple fingers, hands and users as well as for constrained and unconstrained objects.
- The introduction of the concept of Normal Proxies, which extend complex geometric objects with a set of suitable normals for improved grasp detection.
- Plausible object behavior with respect to the users' hand motions is generated with a pseudo-physical simulation.
- Expert interviews, carried out to identify the potential of our approach, show that functional analyses of constrained and unconstrained objects in the car interior can be easily performed using our interaction metaphors in a CAVE.
- Display-related differences of our interaction metaphors are quantified by the comparison of task performance and preferences in a CAVE-like vs. a HMD-environment. Surprisingly, objective and subjective results were similar for both display types.
- We investigated the influence of a virtual hand representation on user performance in a CAVE and found a general preference for working without a virtual hand if the manipulation does not require a very precise judgment of the hand position in relation to a virtual object.

This article is an extended version of a previous publication on our Natural Interaction Metaphor [19]. It expands the original work by user studies evaluating the influence of the choice of display and the presence of a virtual hand on finger-based direct interaction.

Figure 1: A typical Natural Interaction application with a user interacting with a virtual mirror in a CAVE.

2 Related Work

Interaction with virtual environments (VEs) has been a central research field within the virtual reality (VR) community since its inception. A good interface to the virtual world is one of the key aspects for success in implementing new methods and applications into industrial processes. Bowman et al. classify interaction with VEs in regards to selection, manipulation, navigation, symbolic input and system control [3]. For virtual analyses of human-car interaction, selection and manipulation prove most relevant considering users have to deal directly with parts of the car.

In VR, two types of user interaction with virtual objects are characterized: indirect and direct interaction. Indirect interaction is commonly used in industrial applications. An input device controls a cursor, which is then used to manipulate the scene. Usually objects are selected by ray-based picking. Indirect interaction has the advantage in that an input device offers clear feedback of the interaction state and its changes. The disadvantage is that this technique does not allow an analysis of the corresponding real-world interaction itself. Wherever human-car interaction is the focus (e.g. in the analysis of ergonomics issues), direct manual manipulation is the only choice.

In direct interaction, the hands and fingers of the users have to be transferred into the virtual scene (Figure 1). Therefore, glove-based input devices combined with a tracking system or vision-based approaches should be used. We prefer using optical systems since they are able to precisely track the users' hand and finger motions. In particular, the finger tips are important considering that this is where the interaction takes place.

Since we do not employ haptic feedback the users' hands cannot be prevented from penetrating virtual objects. Burns et al. have shown that users more easily detect visual interpenetration than the visual-proprioceptive discrepancy that is introduced if a virtual hand representation is prevented from penetrating an object [20]. In their study they utilized an HMD and did not consider interactive objects. It remains unclear if their findings apply to projection-based environments as well where users are easily able to visually detect a dislocation of the virtual hand from the real hand. Moreover, it is not always obvious how a non-penetrating virtual hand should behave during complex interaction processes, when the users' hands are grasping through objects. We decided to use a co-located virtual hand to avoid a conflicting behavior of the real and virtual hand.

With direct interaction, it is difficult to explicitly select and deselect objects since input devices are not available. Thus we have to use gestures or proximity information which makes selection more difficult for direct interaction techniques. The manipulation of a selected object has to be realistic, meaning that user input is translated into plausible object motions.

Direct interaction techniques can be further distinguished into pseudo-physical interaction and a physical simulation-based interaction. Pseudo-physical interaction combines grasping heuristics with plausible object motion calculated as a reaction to the input of the grasping hand. Grasping heuristics is a very popular research field with applications in robotics, character animation and user interaction. A variety of grasp classifications and detection algorithms can be found in literature (i.e. [6], [11]]). In our approach, heuristics with very simple rules are used covering all necessary types of grasps for our applications. Comparable rules were used by [8]. They described a metaphor to interact with objects in a car interior on a pseudo-physical basis. The way users interact with objects in the car was carefully analyzed. Simple grasping heuristics and finger motion were used to calculate a plausible reaction of manipulated objects. However, only constrained objects could be manipulated by one hand of one user. This is not enough for the realistic representation of human-car interactions.

Holz et al. described an approach that enables one-handed interactions with freely moveable objects on a geometrical basis [5]. Here, all colliding fingers are defined as grasp pairs. In each frame the motion of the most significant finger pair is applied to the object, which in turn reproduces very realistic object behavior. For this interaction continuous finger-object collisions are essential. In our applications mainly constrained objects are relevant. Continuous finger-object collisions cannot be guaranteed for these objects because they are not free to follow the fingers' motion. For interaction only the most significant pair of phalangal contacts was taken into account, whereas multi-hand or multi-user interaction was neglected. Providing an interaction with multiple fingers (and hands and users) is necessary for our applications.

Physical simulation of direct interactions is the ultimate solution given that hand and finger manipulations of objects are governed by physical laws. Systems for rigid body dynamics, which are capable of simulating object-finger interactions, can be stable and fast. Baraff offers a comprehensive introduction to this topic [1]. Rigid Body Simulations are widely used for game physics and are optimized for a number of objects interacting with each other. To achieve this in a stable and robust manner, it is the case that constraints of the objects have to be given up. Their use in user interaction has been demonstrated (refer to [4]), wherein object reaction on indirect user input was calculated with the help of a physics engine. Borst and Indugula showed a very realistic direct interaction of one user's hand with a freely-moveable object using a commercially available physics simulation [2]. A physical representation of the virtual hand was attached to the tracked hand by springs. The interaction between the hand representation and the virtual object was computed by the physics engine. Typical object constraints of our scenarios are usually provided by these simulations. Multi-user support comes for free as well, since all human input has to be handled by interfering forces applied to objects.

For realistic grasping of objects, friction must also be simulated. The physical simulation of deformable objects in combination with friction simulation could provide the most realistic finger-object interaction, but these simulations are not yet fast nor robust enough for complex manipulations. In general, physical simulations usually have certain limitations with respect to the types of objects (e.g. only convex objects are supported). Our scenarios include arbitrary objects, such as thin sheets and very complex shapes. Unfortunately, the integration of all aspects of Natural Interaction into a single application which is usable in the analysis of ergonomic issues in virtual cars has not yet been realized.

For evaluating our interaction technique we perform an expert review judging the usability of our metaphors. Expert reviews usually reveal 75 percent of all usability problems with the help of only five experts [9]. With the developed metaphors we focus on applications using CAVE-like VR-systems. However, the automotive industry also employs Virtual Seating Bucks which use HMDs for virtual assessments. Although extensive work has been done in the fields of display technology and interaction metaphors only few studies exist on the impact of display choice on interaction. Comparing quantitative measures and task performance in several displays [16], [17] and [18] report display-related differences in individual task performance and user satisfaction for search tasks and scientific data exploration. In [15] virtual hand and virtual pointer techniques where evaluated in a CAVE-like environment and a panoramic display. All studies identify display properties such as field of view, field of regard and image quality as a source for differences in task performance and subjective judgments. Consequently we also expect differences for the usability of our interaction metaphors in the CAVE vs. an HMD. The CAVE provides a higher field of view and it is generally more accepted by our users. On the other hand the HMD does not suffer from focus and convergence problems if direct hand-based interaction is required.

3 Grasping Heuristics

The ability to grasp objects is essential for the realization of Natural Interaction in virtual environments. The moment of establishing a valid grasp is characterized as the selection part of the metaphor. While grasping, the users' hand movements manipulate the grabbed object by transforming it. The end of interaction is defined by deselecting objects through releasing the grasp.

As previously mentioned, grasping has been extensively studied within several research fields. One way to detect human grasp is by using Grasping Heuristics. Here, valid grasps are detected by rules or conditions that define the beginning and the end of a particular grasp. Two parties are involved in a grasp: the virtual objects and the users. Grasping conditions are derived from the relationship of one party to the other. Therefore it is necessary to have a virtual representation of at least the users' hands in the VE. The basis of our heuristics is the detection of collisions between hand and object geometries and their orientation to one another.

The definition of start conditions is subject to a tradeoff. On one hand, every single grasp intention of users has to be detected. On the other hand, unintended grasps have to be avoided as much as possible. Both requirements cannot be offended without annoying the users. The same is true for stop conditions. They have to provide stable grasping despite unintended hand movements, tracking jitter or even tracking interruptions. At the same time, it is unacceptable if a dropped object were to stick to the users' hands.

3.1 Grasping Conditions

Our grasping heuristics are triggered by collisions between the finger phalanges and virtual objects. These collisions are detected on a per-triangle basis by our VR-Software. Whenever at least two finger phalanges collide with a single virtual object, we check if both phalanges establish a grasping pair with respect to this object. Valid grasping pairs are finger phalanges that have a virtual object between them. They are valid if the rules of a pseudo-physical replacement of friction are satisfied. This means that physically correct friction is approximated by a geometrical representation called friction cones, as suggested by [5].

To explain our friction cone method, we must first define several terms. A *collision pair* consists of a finger phalanx and a virtual object that are colliding (in contrast to a *grasping pair* that consists of two collision pairs clamping the very same virtual object). The calculation of the collision point and the collision normal considers all triangles of the object's surface, which are involved in the collision. We define the *collision point* as the center of all vertices of these triangles. The *collision normal* is calculated by averaging all involved face normals. In [5] all face normals within a certain radius of the collision center are collected and averaged. Their approach smoothes rough surfaces and makes them graspable.

However, we have to find another way of approximating normals since we often have to deal with several box-shaped objects, including sun visors, the interior mirror or items that have to be placed in storage compartments such as books. Averaging normals of perpendicular faces leads to unusable collision normals that could prevent valid grasp pairs. The very same problem occurs while interacting with tiny objects that are not significantly bigger than the geometry of the finger tip. Here, a collision with the object can also include perpendicular or opposing faces resulting in unusable normals if the averaging approach is applied. To avoid distortions of collision normals and the resulting problems with our grasping heuristics, we use appropriate grasping proxies that replace rough surfaces or perpendicular faces. These proxies are generated such that they provide normals that are compatible with our grasping heuristics. Details are explained further in section 4.

In order to calculate the friction cone and to decide if two collision pairs define a grasping pair, the collision points are connected by a line. The angle between the collision normal and the extended line is the friction angle (Figure 2). The smaller the angle, the stronger the force that finger phalanges can apply to the object and thus the tighter the object can be clamped by the grasping pair. If the friction angles of both collision pairs are smaller than a pre-defined threshold, these collision pairs define a valid grasping pair. Surfaces with different friction values can be simulated by using friction cones with different angles. High thresholds simulate rough or even sticky surfaces, whereas small angles are used for smooth or slippery objects.

An object is defined as being grasped if at least one valid grasping pair exists. A grasp is defined by a virtual object and all valid grasping pairs related to this object. It is irrelevant which hand the collision pair belongs to. Using this definition, multi-hand and even multi-user grasps can be uniformly treated.

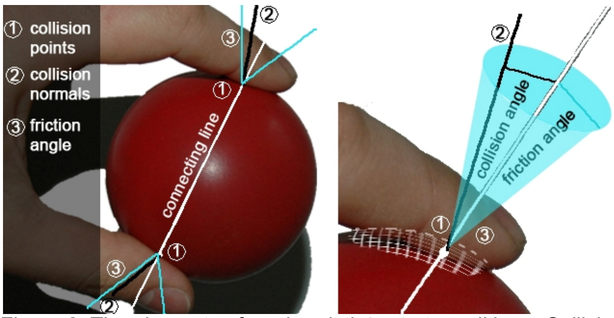

Figure 2: The elements of our heuristic´s start conditions: Collision Points and Normals and the Friction Cone of a grasping pair

3.2 Stop Conditions

A grasp of an object is valid if at least one valid grasping pair is detected. In each application frame, the virtual environment is checked for newly valid grasping pairs that need to be added to their associated grasp. Each existing grasping pair is checked if it is still valid with respect to the stop conditions. A grasp is valid as long as at least one valid grasping pair is associated with it. We define two stop conditions responsible for detecting disappeared grasping pairs.

The first stop condition compares the initial distance between the collision points of a newly detected grasping pair to the distance of the corresponding finger phalanges during interaction. A grasping pair is no longer valid if the current distance increased more than a user-defined threshold from the initial distance. This stop condition aims at mimicking the users' intention to drop an object by opening their hand (Figure 3). The threshold is necessary to compensate for jitter and the imprecision of users trying to maintain a particular grasp. The calculation of the distance of the phalanges' collision points requires that the initial collision points be stored with respect to the local phalanx coordinate system of the involved phalanx. In each frame the current pose of the involved phalanges is used to transform the stored collision points to their current position, which is needed for the evaluation of the stop condition. We call these updated collision points the *contact point* of a collision pair.

The second stop condition considers the users' intention to drop an object by moving their involved finger phalanges away from the grasped object. This is an important condition if constrained objects are involved, since they cannot always follow the hand movement. To better explain this condition, we define the *barycenter* of a grasping pair to be the mean of both of its contact points. We store the initial distance of the grasping pair's barycenter to the object center. This distance is continuously compared to the distance between the current barycenter and the object center. The grasping pair is valid as long as the current distance is smaller than the stored distance times a user-defined threshold. In most cases it is reasonable to assume that the range of finger movements is small for small objects and grows commensurate with the size of an object, considering users tend to perform tiny and careful movements when dealing with small objects and coarser manipulations when dealing with larger objects.

If one of both stop conditions is fulfilled, the corresponding grasping pair is deleted. If no valid grasping pair remains for a grasp, the grasp is deleted as well. During interaction it is possible that grasping pairs appear and disappear. Our stop conditions differ significantly from those described in [5] since they are not based on continuous collisions of the finger phalanges. Constrained objects cannot freely follow the movements of the finger phalanges. Thus, it cannot be assumed that the collision of the finger phalanges and the virtual objects can be continuously preserved during interaction. Instead, our stop conditions are heuristics, which try to match the users' intention to release an object.

3.3 User Input

During a valid grasp – from the moment the first grasping pair is detected to the moment the last grasping pair is deleted – the users are applying motions to the grasped objects. In reality, the forces applied to an object are responsible for its pose change. Since we do not use

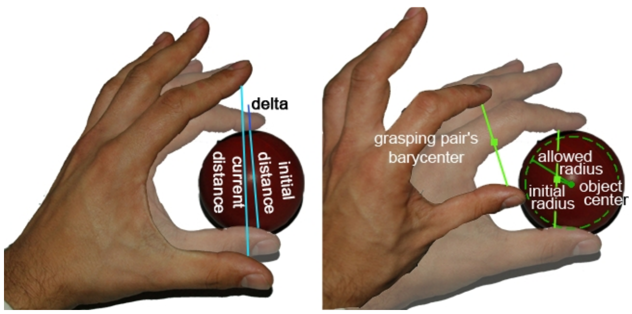

Figure 3: The stop conditions of our heuristics: Contact point distance (left) and the grasping pair´s distance to the object center (right)

physics simulations, there has to be a plausible object motion derived from the motion of the finger phalanges.

For each grasping pair we basically calculate a frame-to-frame contribution based on the involved phalanges' movement and then average these contributions to compute the object motion. The object translation – the so called *way* – is simply derived from the averaged movement of the barycenter of each involved grasping pair. For the rotation contribution of each collision pair, we take three different rotations into account. The main rotation (see Rotation 3 in Figure 4) is derived from the movement of the line connecting the two involved phalanges. In this case, the rotation results from a translation of the fingers. Additionally, the rotation of each single phalanx has an effect on the object (see Rotation 1 and 2 in Figure 4). The rotations of the phalanges can be directly calculated from the orientation change of each involved finger phalanx. These three rotations are averaged and constitute the rotational contribution of a grasping pair.

All translational and rotational contributions of all grasping pairs belonging to a grasp are averaged and applied to the grasped object. The *grasp center,* which is also the center for object rotation, is defined as the mean of the barycenters of all involved grasping pairs.

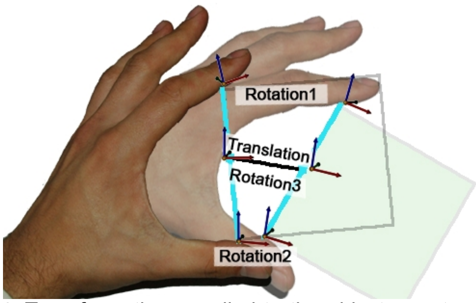

Figure 4: Transformations applied to the object: one translation and three averaged rotations

3.4 Push Input

Not every object in the car interior needs to be grasped for manipulation. Some objects, especially flaps and buttons, are usually pushed by one or more fingers or by the palm. This means that they are not clamped between the fingers but instead they back off from the finger phalanges in case of a collision while respecting their constraints. Since a pushable object always responds to any finger collision, this would impede robust grasping. Consequently, we define objects then to be either pushable or graspable. However, a combination of push and grasp input must be possible for large objects. In this case, parts of an object can be defined to respond to push input and other parts can be defined to be graspable. Object reaction is then applied to the entire object.

To enable this intuitive interaction, we need to compute a plausible object motion resulting from pushing fingers. The translational contribution of each involved collision pair is derived from the frame-to-frame distance that lead to a collision of the phalanx with the object. Rotational input is ignored. The contributions of all collision pairs are averaged and applied to the object as is done with the grasping pairs for graspable objects.

This simple penalty method leads to object reactions that tend to be slightly exaggerated, since we do not calculate the exact finger penetration into the object. However, this easily computable approach provides a convincing object motion because the back-off motion from an object is only minimally incorrect due to the limited hand velocity. We do not allow freely moveable objects to be pushed since their reaction on push input cannot sufficiently be simulated using our method.

4 GRASPING PROXIES

The surface normals of objects are used to calculate friction cones, which determine the validity of a grasping pair. It is essential for robust grasp detection that the collision normal is a reasonable representation of the surface touched by the users. Collision normals are commonly calculated based on the surface normals of the faces, involved in the collision with the finger phalanges [5], [8].

There are three problems with the direct use of these collision normals in regards to the robustness of our grasping heuristic. First, there is the well-known problem of real world CAD models that still lacks a solution: flipped normals may occur for individual faces. Second, the collision of a finger with a box-shaped object (e.g. the interior mirror) can lead to a distorted collision normal due to the involvement of perpendicular faces (Figure 5). A similar problem occurs while interacting with an object that is as small as the finger geometry. It is almost impossible for the users to grasp such objects in a way that will result in a valid grasping pair, which requires that two fingers are colliding with faces having normals with almost opposite directions. Instead, for small objects it often happens that a single finger phalanx collides with a set of triangles that have perpendicular or even opposing normals. Third, our scenarios deal with complex geometry, which results in a high collision detection effort and may lead to decreased frame rates. However, this last problem has become less important due to efficient collision detection algorithms and faster processors.

A common solution for these problems is to substitute inappropriate geometry with simpler collision proxies. These substitutes are often used to approximate geometry in physically-based interaction scenarios [4]. They can also be used with our approach to define appropriate normals for grasp detection. These simplified proxy geometries are only used for grasp detection and the computation of the physically realistic object motion, whereas the object is still rendered from its original representation.

This approach performs the collision detection through a simplification of the original object geometry. While this is an advantage concerning speed and grasp detection robustness, it introduces an error due to the simplification involved. We therefore suggest using the concept of *normal proxies* instead of collision proxies. Normal proxies extend the object description by additional normals, which are only used for grasp detection. The original normal responsible for lighting is not changed. With this approach, the collision detection still refers to the original geometry

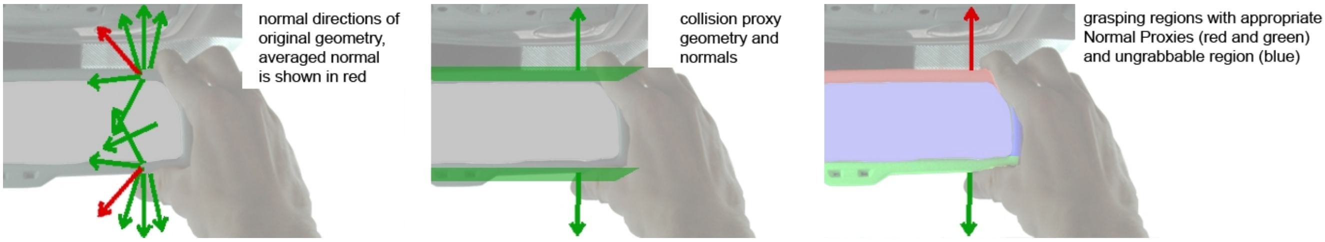

Figure 5: Averaging contact normals causes problems with the original geometry (left), collision proxies (middle) and normal proxies (right) provide reasonable normals

but for grasp detection, an appropriate proxy normal can be used. This allows us to overcome the annoying flipped normals problem, but more importantly, we can overcome the problem of averaging unsuitable normals from a neighborhood of a colliding phalanx.

One challenge with normal proxies lies within data preparation. A normal, providing robust grasp detection, has to be assigned to each face of the object. The direction of the normal proxy depends on the location of the face on the object's surface and the way the user is grasping this particular object. In our scenarios, usually a small number of varying normals are necessary per object. All objects are designed to be used in a defined way and can be reduced to a small number of grasping surfaces or areas with a common normal proxy. To keep the data preparation effort low, we use a semi-automatic approach to define the normal proxies. We divide an object into grasp regions which correspond to the areas where the fingers are typically located during a grasp. Each grasp region contains a set of faces and has a common normal proxy. Additionally, to avoid unintended grasps we introduce regions where an object cannot be grasped.

The concept of these different grasp regions is explained using the interior mirror as an example shown in Figure 5. The interior mirror has three grasp regions. The first region has a grasp normal that is pointing upwards and contains all triangles of the upper part of the mirror. The second region groups all faces of the lower part with a normal pointing downwards. The third grasp region contains all remaining faces and is defined as not graspable. This subdivision is due to the fact that the interior mirror is usually clamped between the index finger and thumb on its upper and lower sides. Without a proxy method, the averaging of normals of colliding phalanges may make it impossible to actually grasp the mirror. Our normal proxy approach allows us to perform the collision detection on the original geometry, but the grasping heuristics is provided with normals that provide a robust grasp (Figure 5).

All three approaches (using the complete original geometry, using collision proxies or using normal proxies) are compatible with our grasping heuristics. Collision proxies are fast and robust but inaccurate with respect to the actual collision. Using the original geometry may make it impossible to grasp an object. Normal proxies are a good compromise for increasing the robustness of grasping heuristics. The disadvantage is that they cannot be easily constructed by a fully automatic approach. However, since our car interiors are typically constructed from a set of defined building blocks, we simply need to create the normal proxies only once and then store them with the objects in the building block library.

5 Plausible Object Reaction

In reality, human-object interaction is a result of physical processes. Basically these processes can be simulated for user interactions with virtual objects, but as we explained before, until now the complex scenarios found in our applications cannot be completely simulated. One aspect of the physical processes is that the users apply forces to the objects with their hands. We explained in section 3 how our metaphor translates these processes into our pseudo-physical model. The object's behavior is a reaction based on this input and, therefore, follows physical rules as well. Forces applied by the users interact with the inner forces of the object (e.g. Gravity, Friction and Inertia) and cause it to move. A further influence on object motion can be caused by restrictions of its degrees of freedom (DOF). An object has a total of six DOF if it is not restricted (3 DOF of translation and 3 DOF of rotation). If an object has six DOF, the force that is applied to the object results in a motion in the direction of the superpositioned input forces (e.g. user input, Gravity, etc.). For constrained objects having less than six degrees of freedom, the objects' reactions are more difficult to compute.

First, we describe how objects that are not restricted react toward user input while being grasped. Then we discuss the case regarding restricted objects, wherein constraints can occur and how the objects' motion is influenced by these constraints.

5.1 Unconstrained Objects

As mentioned before, unrestricted objects have six degrees of freedom. In reality, freely moveable objects that are not grabbed by the users would fall due to gravity until they would hit the floor, for example. The simulation of gravity is not necessary in our scenarios. Objects that can be grabbed and moved through space without any restrictions nevertheless are necessary (e.g. for clearance analyses). If a non-restricted object is grasped in reality it follows the user-induced forces instead of secondary forces including gravity and air resistance.

The object's reaction to the user input introduced by a grasp can be easily calculated from the rotations and translations of the corresponding grasping pairs. We calculate the mean of their rotations and the grasp center for the current rendering frame and the previous frame. The grasp center is the center of rotation defined by the grasping pairs. The rotation is then calculated as described in section 3 and the translation is defined by the frame-to-frame movement of the grasp center.

Our method of pseudo-physical behavior addresses both, coarse motions defined by long distance movements of the whole hand and fine-grained motions as they are intended by tiny shifts of the two phalanges belonging to one grasping pair.

5.2 Constrained Objects

Freely moveable objects move as expected by the user. However, most objects found in our scenarios have restricted degrees of freedom since they are mounted to the car body in one way or another.. We therefore define three different types of constraints:

Mounting constraints are constraints that are caused by the way objects are mounted onto the car body. An example includes the glove box lid mounted to the cockpit by a hinge. These constraints are well known from the field of Rigid Body Dynamics.

Functional constraints further influence object motion within a mounting constraint and represent a certain function of the object (e.g. lock positions of controllers).

Location constraints result from the clearance of an object. They are caused by collisions with other objects. Constraints of this type are often avoided by functional constraints. In this case, engineers restrict the objects' motion such that they do not collide with other objects.

Each type of constraint has a certain influence on the objects' reaction to user input. The following sections explain how this influence is addressed by our metaphor.

5.2.1 Mounting Constraints

Mounting constraints are well known from Rigid Body Dynamics, where they are generally considered as joints linking objects to each other. Joints are mainly characterized by the number of remaining degrees of freedom they provide. A list of mounting constraint types, provided DOF (R describes rotational and T translational DOF) and examples are given in Table 1.

TABLE 1: TYPICAL MOUNTING CONSTRAINT TYPES

Type	DOF	Example
Guide Plate	1T	ESP-Button
Pivot Joint	1R	Light Controller
Hinge	1R	Glove Box Lid
Ball and Socket Joint	3R	Interior Mirror

Additionally there are objects that follow kinematic chains of two or more joints (e.g. the turning light lever is restricted to two degrees of freedom rotations by two hinge joints in a row). Kinematics that can be found in our scenarios are usually quite simple. Another example is the steering wheel adjustment that is defined by a combination of a hinge and a guide plate resulting in a 1R1T constraint. These simple kinematic chains can be represented by the hierarchical structure of the scene graph. Each joint in the chain can be modeled by a matrix node arranged in a hierarchy similar to the kinematic chain. Object reaction with respect to user input is calculated and separately applied for each of the joints. The interaction object is child to all group nodes and experiences a combination of all separate transformations.

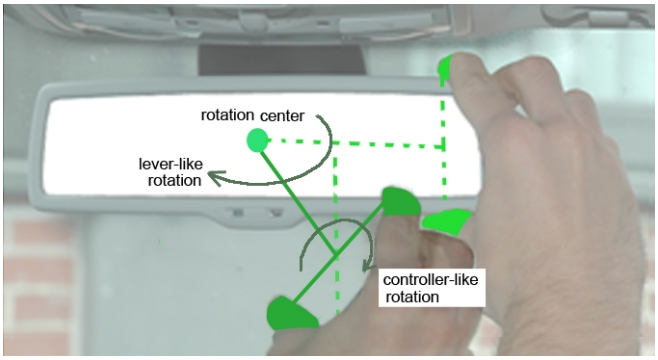

Figure 6: Rotation interaction: The combination of lever-like and controller-like interaction results in a correct rotation of the 3R-restricted object

To consider the influence of the constraints, the dimensions of the user input have to be reduced with respect to the allowed degrees of freedom. For objects that are not allowed to rotate this means to only take the displacement of the grasp center into account, rotation can be completely ignored. For 1T- and 2T-restricted objects this can be easily achieved by projecting the motion vector onto the constraint vector or plane.
Rotational constraints are slightly more complicated to consider. It is necessary to understand how users interact with such objects to compute a plausible object reaction. There are basically two different ways of interaction with rotationally restricted objects. A *lever-like rotation* occurs if the grasp center has an offset to the rotation center of the object. The user rotates the object by moving the grasp center around the rotation center. One-handed turning of the steering wheel is an example of this kind of interaction.

Small controllers have to be handled differently. Here the users apply a rotation by turning the wrist of the hand. The grasp center then equals the rotation center of the object or lies close to the rotation axis. We call this rotation *controller-like rotation.* To derive user input from the translation of the grasp center would cause an incorrect object reaction in this case. Therefore it is crucial to use the rotation of the grasping pairs instead.

For large objects, both types of rotation may have to be considered for the final object reaction. For example, the two-handed turning of the steering wheel usually is a combination of lever-like rotations of grasping pairs composed of phalanges of one hand and controller-like rotations of inter-hand grasping pairs (see Figure 6). The same is true for 3R-restricted objects. Here, the object rotation is a combination of a lever-like rotation around the rotation center of the object and a controller-like rotation around the axis defined by the rotation center and the grasp center.

All possible interaction scenarios can be covered if the rotation constraints are considered at the grasping pair level. For 3R-restricted objects, both rotation types are calculated for each grasping pair. For the grasping pairs of any other rotationally restricted object, we have to consider two separate cases. A lever-like rotation is calculated if the grasping pair's barycenter is far away from the rotation axis and the rotation center. For grasping pairs with barycenters lying on or near the rotation axis or rotation center, we use the controller-like rotation. Allowing rotations of the other type would lead to incorrect results in each of the respective cases.

The lever-like rotation is derived from the translation of the grasping pair's barycenter around the rotation center as it is suggested by [8]:

- Calculate the vectors between the object's rotation center and the grasping pair's barycenter for the current frame (*RBC*) and the previous frame (*RBL*).
- For restricted objects project *RBC* and *RBL* into the constraint plane defined by the rotation center and the rotation axis of the object.
- Calculate the rotation axis as the Cross Product of *RBL* and *RBC.*
- Calculate the rotation angle from the Dot Product of *RBL* and *RBC.*

Controller-like rotations are calculated from the inter-frame difference of the line connecting the two collision centers of the grasping pairs. This rotation is part of the rotational motion intention introduced by the grasping pair as explained in section 3. Here, this rotation has to be reduced to a rotation around the rotation axis in the very same way as it was done for the lever-like rotation. This is done by projecting the two vectors into the plane perpendicular to the rotation axis. Other rotational influences of the grasping pair can be neglected since they cannot introduce a rotation around the axis. For 3R-restricted objects, the rotation axis equals the line connecting the rotation center and the barycenter of the grasping pair.

The individual rotation quaternion of each grasping pair can be merged to compute the total rotational influence of a grasp by using the SLERP algorithm [10]. With the constraint-based modification of each grasping pair's influence, the overall grasp rotation respects the object's constraints and covers all described types of rotations.

5.2.2 Functional Constraints

In the previous section we explained how the objects' motions are constrained by the way they are mounted to the car body. There are further constraints that are constructed by engineers to realize a certain function of the object. For example, a controller that can be rotated around one axis can be further restricted by lock positions at which the states of the controller are switched. Stop positions limit the object's movement beyond a certain point. Consequently, functional constraint cannot be found with freely moveable objects.

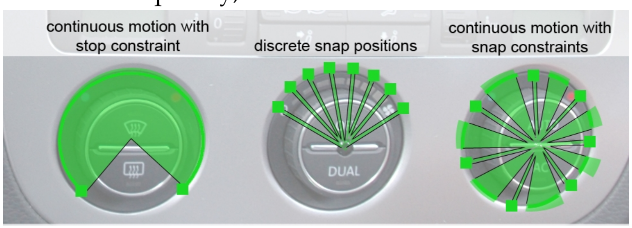

Figure 7: Three examples for functional constraints

Lock position constraints can have several variants. The objects can be allowed to perform continuous motions with states that are set at defined positions without any influence on the object's motion. Another possibility is that these positions are realized using a snap function. Finally, the third option is that the motion of the object is discrete, allowing the object to snap only into the lock positions. A stop position constraint is defined by a range the object is allowed to move within (Figure 7).

For the consideration of functional constraints, we introduce a ghost object, which moves as if no functional constraints apply. The offset of the current pose of the ghost object to a valid design position – a 3D vector for translational and an angle for rotational objects – can be used to check for violations of functional constraints and for their definition. The offset can be calculated for each frame or continuously tracked during interaction. The current offset of the ghost object is compared to the predefined positions in each consecutive frame. If the current offset equals a lock position, it is reached. For objects with a snap function, the constraint is extended to a snapping range, whereas for objects moving discretely, all snapping ranges have to adjoin each other.

As long as the ghost object is not influenced by constraint condition, its pose is transferred directly to the pose of the actually manipulated object. Under constraint condition, the object is set to the lock position. For stop constraints, the pose transfer is only performed within the allowed range.

Since the object stays in lock position until the users' input moves the object out of the Epsilon area, a difference between users' motion, finger motions and the actual object motion can only be observed if it is intended – in the snapping or stopping case. If the user is releasing an object when it is under constraint condition, the ghost object has to be set to the pose of the rendered representation to avoid inconsistencies.

5.2.3 Location Constraints

Location constraints are constraints that are caused by object-object interaction. An interactive object collides with another object – not necessarily an interactive one – and this collision works as a constraint for the object motion. These location constraints represent a limit for our Natural Interaction metaphor. In handling object-object interaction, the use of physical simulation is required. The processes resulting from object-object collisions are too complex to represent them on a pseudo-physical basis.

However, for most of such situations in our applications, a plausible reaction to this constraint type is possible. The collision of a restricted interactive object with a static part of the car body usually causes the interactive object to stop. For example, the sun-shield colliding with the windshield stops its motion immediately, as well as the interior mirror colliding with the windshield. For these situations we can transform location constraints into the functional constraint type of a stop position (see 5.2.2).

The plausible reaction of a freely moveable object colliding with a static part cannot be realized using our metaphor. However, we need such scenarios for clearance analyses. In order to get around this, we visualize such collisions by highlighting the involved triangles. The interaction between two or more interactive objects is also not within the scope of our metaphor. Therefore the use of physical simulations is required.

6 Expert Review

We carefully evaluated the Natural Interaction metaphors. As an initial study we tested for usability and realism of our approach. We performed an expert review since there is no quantitative measure for these issues.

6.1 Methods

Participants

We performed the review with six experts for automotive VR applications. They were ergonomists, simulation experts and automotive VR system developers.

Apparatus

The expert review was conducted in our three-sided CAVE, driven by a PC-Cluster generating images for 10 full-HD projectors. A pair of two projectors provides a passive stereo image separated by the INFITEC technology. For each of the side walls two pairs of projectors are vertically arranged with horizontal edge blending resulting in a total resolution of 1920x1920. The floor image is diagonally projected by one projector-pair from above. Precise finger tracking is realized through the wireless finger tracking gloves from Advanced Realtime Tracking. Optical hand and head tracking is supported by the same camera system. As a test scenario, we chose the interior of a Volkswagen Touareg consisting of approximately two million triangles, running at 12-15 fps.

Design and Procedure

We performed semi-structured interviews with three pairs of experts to allow for a free discussion between the interviewees. Each session started with a testing period of approximately 20 minutes for the pair of experts. We asked them to interact with all objects during this period. During interaction, the experts already discussed the prototype. After each expert experimented with our interaction metaphors, we interviewed them together for another 20 minutes.

6.2 Results and Discussion

Judging the actual grasping, our experts stated that they were able to select objects very intuitively and robustly with the help of our grasping heuristics. Problems occurred with smaller objects due to occlusion. Interestingly it turned out that grasping the objects with a tip pinch worked best for the users considering it provides least occlusion and the most precise way of interaction.

Concerning the release of an object, the experts considered it mandatory to carefully balance the stop condition parameters. Sometimes objects appeared to be sticky and sometimes objects tended to follow the finger motion even though the users already opened their hand with the intention to end the grasp. The threshold of the first stop condition (see 3.2) was responsible for both of these effects. However, a small threshold can result in unstable grasp continuity due to jitter and imprecise finger motion. This trade-off could be solved by a calibration of this parameter or a careful adjustment per object. However, experts became quickly proficient with the technique and precisely released objects by dropping them carefully. Furthermore, the second stop condition was appreciated as a definite interaction stop in cases where objects appeared to be sticky.

The motion of the objects as a reaction to finger collisions was judged to be very plausible and realistic. In particular, the precise response to very fine grained motion of the finger tips fascinated the experts and was highly recommended. Multi-hand interaction was successfully tested by passing the bottle from one hand to the other. The same would be possible with our method for two or more users in a multi-user stereo setup.

The push interaction was also judged to be very intuitive since object reaction immediately follows a collision. However, the experts complained about the missing mass and inertia of the objects. At first sight it was unrealistic for them that objects did not move according to their size and mass when touched by the user.

With our direct interaction metaphors, the experts were able to judge the accessibility of objects. In contrast to the indirect method they were used to – involving an input device such as the Flystick – they were able to select the objects directly with their hands. Moreover, they found our method to be more intuitive since no button assignment needed to be learned. Due to the direct interaction involving the hands and arms of the user, our experts identified a potential range of new applications for our techniques. This range included assembly simulations which need to consider the clearance of the human hand – especially concerning car maintenance – to the assessment of ergonomics issues in the car interior. Furthermore, the experts stated that this interaction would enrich immersive design and concept reviews due to the increased interactivity and immersion, enabling a more realistic experience of the virtual model. However, our object selection and release method was stated to be less precise than pressing and releasing the button of an input device.

An interesting question raised was if the experts would prefer to use a virtual representation of their own hands during interaction. Despite being occluded by the real hands, the virtual hands can be seen by the users since our simple hand model did not perfectly match the real hands. Half of the experts preferred the visual feedback giving them the opportunity to judge their virtual fingers' relation to the objects. The other half preferred to interact without a hand representation. They appreciated that they were able to concentrate on the virtual object without being distracted by the virtual hand representation. They further noticed that because they were not tempted by the virtual hand model to look at their real hands, they had fewer problems with the focus and convergence mismatch. Since the expert review did not reveal a clear preference we decided to further investigate user preference and the influence of the virtual hand on grasping robustness.

Finally, our experts remarked that for the assessment of object clearance in very complex maintenance and assembly scenarios object-object interactions through collisions are required.

7 Natural Interaction in a CAVE vs. an HMD

At Volkswagen, we mainly use projection-based immersive environments for the functional evaluation of car concepts because of their good user acceptance due to the large field of view and decent comfort. However, we also have applications which require the use of HMDs. Virtual Seating Bucks are such an example, where the virtual interior of a car is registered with a basic car mockup consisting of a real car seat, steering wheel, etc. [12]. Seating bucks have the advantage that functional assessments benefit from the passive haptic feedback of the real-world components. We believe that HMD-based applications of this kind also benefit from our metaphors. Consequently, we investigated the impact of the display type on the usability of our Natural Interaction metaphors.

7.1 Methods

Participants

Twelve subjects participated in our assessment, one being female. The age ranged from 22 to 39, two of them were left the rest right-handed. All subjects had unimpaired or at least corrected sight. All of them were already used to the interaction metaphors.

Apparatus

We conducted this experiment in the same interaction scenario that was already used for the expert review. For the CAVE-condition we used the three sided CAVE that was described in section 7.1. For the HMD-condition we used the same CAVE-environment, but the virtual environment was provided by a Rockwell Collins SR80 Head Mounted Display having a diagonal FOV of 80° with a resolution of 1280x1024.

Design

A common measure for the evaluation of interaction metaphors and their influence factors is to compare task completion times (TCT), as suggested by [14]. Direct interaction tasks can be separated into three phases: selection, manipulation and deselection. In the context of natural interaction metaphors, the phases are realized by the reliable grasping heuristics, a realistic and intuitive manipulation and a reliable release of the object. The TCTs directly depend on the usability of the implementation of each task phase. Unfortunately the TCTs are influenced by many factors, such as tracking reliability and individual task performance. Subjective judgments and the evaluation of user preferences help to complement the performance information provided by the TCTs.

For the TCT-measurement reference points for each manipulated object and target positions for these points were defined. The target position of an object is reached if each of its reference points is within a tolerance region of the respective target reference points (Table 2). The reference points were visualized by small red spheres. The target position was shown with the help of semi-transparent colored duplicates of the objects and the tolerance regions were visualized by semi-transparent green spheres with of appropriate radius (Figure 8).

Figure 8: Sun shield interaction task with the target visualization

Table 2: Control Points and tolerance of the objects

Object	N° of reference points	Tolerance (mm)
Sun Shield	1	20
Driver's Door	1	30
Interior Mirror	4	10
Soda Bottle	2	20
Light Switch	1	5

Once the object reached its target position, a green sphere appeared around the virtual car. As the start and stop condition for our measurements, we defined the presence of the index finger tip within a semi-transparent purple sphere around the center of the steering wheel (r = 80mm) that became opaque when the condition was reached. Since there are objects that are manipulated with the left hand (door, light switch) or the right hand (sun shield, mirror, bottle) we chose the corresponding index finger for the start-stop decision.

Users had to perform each task three times. The objects were reset to design position each time. The subjects were told to carefully place the objects at the target position. However to fulfill the task it was sufficient that the object reached the target position once for a short time. It was not necessary that it stayed in target position after releasing to avoid multiple grasp and release sequences.

We asked the subjects to rate the quality of each task completion on a seven-point Likert-scale (7: very good grasping, placing and releasing; 1: very poor) to complement performance measures. For further evaluation we additionally used a short questionnaire presented after all tasks of each condition. Users were asked to rate the grasping, the releasing, the realism of the interaction and the overall ability to judge functional aspects of a car with each system on a seven-point Likert-scale ranging from one (very poor) to seven (very good). In all cases we tested for statistical significance by T-tests with repeated measures.

The reference for the described tasks is the task performance in a real car interior. Therefore we observed a user in the very same real car interacting with the very same objects. Of course the target positions differed slightly and we had to time manually. Users had to touch the badge at the steering wheel for initiating and finishing the interaction. Each interaction was performed three times and TCTs were averaged (Figure 9).

Procedure

The experiment was conducted in the already described scenario used for the expert review. Before each test we calibrated the finger tracking system to the individual users' hands. Half of the group started in the CAVE the other half started wearing the HMD to counterbalance for order effects. In both conditions a virtual hand representation was used. We defined five tasks all composed of grasping an object, moving it to a defined target position and releasing it. The tasks included the rotation of the driver's sun shield from design position to stop position, the opening of the driver's door until it reaches its stop position, the rotation of the interior mirror towards the user, the passing of the soda bottle from the passenger's footwell to the center console and the rotation of the light switch to its stop position. No push interaction was allowed, so each object had to be explicitly grasped. With the chosen objects a broad range of object types typically found in cars is covered including several kinds of constraints and different object shapes, locations and sizes.

7.2 Results and Discussion

User Preference and User Observation

We were interested in the influence of display type on task performance and user preference. Surprisingly the user preference was quite balanced. From our everyday experience we would have expected a strong preference of the CAVE since HMD-applications usually suffer from limited user acceptance. For our test scenario concerning Natural Interaction seven users preferred the Head Mounted Display, while five preferred the CAVE. Since the users had the chance to directly compare both systems they were able to perceive the advantages of Head Mounted Displays over projection-based systems and vice versa. HMDs fully exclude the real world and thus avoid the typical focus and occlusion problems resulting from the interaction between real hands and virtual objects. Furthermore we believe that the HMD application strongly benefits from our Natural Interaction metaphor, because users have a self-reference – their virtual hand – in the virtual world, which has been shown to improve size perception [13]. Thus it is no surprise that users did not complain about incorrect size perception of the car interior. However some of the users criticized that the narrow field of view (80° diagonal) of the HMD made it necessary to extensively move the head during interaction. Although this particular display is relatively comfortable the increased discomfort and fatigue were mentioned by some of the users. Interestingly some users found it difficult to judge the distance of virtual objects in the HMD condition. No user complained about the virtual hands penetrating virtual objects during grasps.

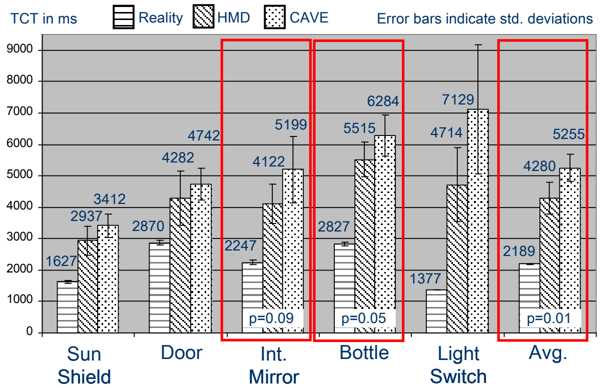

Figure 9: Per-object and average task completion times in reality, HMD and CAVE. Framed columns represent at least weak significant differences of HMD and CAVE.

Task Completion Times

Per-object mean task completion times did not show a significant advantage for any of the conditions. However a tendency towards faster interaction in the HMD condition can be observed if the fastest TCTs for each task compared. Here, a significantly faster interaction was achieved for the sun shield and the bottle ($t=-2.3$, $p=0.04$ and $t=-2.2$, $p=0.05$ respectively, $df=11$). A weak significant effect was encountered for the mirror ($t=-1.8$, $p=0.09$, $df=11$). Average task completion times for the door and the light switch were considerably faster, but did not become significant due to large variance. Over all tasks the average of the fastest interaction times was significantly faster ($t=-3.0$, $p=0.01$, $df=11$) for the HMD-case (Figure 9).

Comparing average TCTs of the virtual scenarios with those measured in reality reveals that virtual task performance is two to more than four times slower. This shows that the users still act more carefully and slowly in the virtual scenario. Although our Natural Interaction metaphors provide robust and intuitive grasping and manipulation there is still potential for improvement (Figure 9), which probably cannot be fully exploited without some sort of haptic feedback.

Subjective Measures

The results of the subjective interaction judgments were comparable in CAVE and HMD condition (Figure 10). For each system one object was judged significantly better – the sun shield interaction under CAVE-condition ($t=-3.0$, $p=0.01$, $df=11$) and the light switch under HMD-condition ($t=2.4$, $p=0.03$, $df=11$). This complies to our expectation on display influence since the sun shield interaction does not require a precise judgment of the hand position in relation to the sun shield. In contrast, the spatial relation of the user's hand to the tiny light switch is easier perceivable in the purely virtual environment of the HMD.

Concerning the short questionnaire again both systems were rated comparably with a tendency for better values when interaction took place in the CAVE (Figure 11). The only significant effect was that releasing objects was rated better in the projection-based system (t=-2.3, p=0.044, df=11), which astonishes since the actual grasping – and releasing – heuristics were the same in both conditions. We were pleased to see, that for both systems interaction realism and the ability to judge functional aspects of a car were said to be above average with rates of 4.58 and 5.08. Surprising for us was that for the subjects the HMD-application was more appropriate for functional car assessments than the CAVE-environment (5.08 and 4.83, t=0.6, p = 0.59, df=11). The advantage of the HMD is not a significant effect but if anything we would have expected a significant effect in favor of the CAVE due to the reduced FOV and comfort of the HMD.

In general our tests show that for functional assessments of the car interior our Natural Interaction can be equally well used in both systems – the CAVE and the HMD. The decision which system should be used can be made based on application-dependent factors, such as the requirements on passive haptics or the size of the manipulated objects.

8 Evaluation of Virtual Hand Representation

Our expert review – performed in the CAVE – revealed no clear preference whether the users' virtual hands should be visualized or not. Half of the experts claimed that a hand representation helps to judge the hands' location with respect to the virtual objects which is a prerequisite for reliable grasping. The other half stated that they were disturbed by the virtual hand models since they tempted them to look at their hands instead of the objects (cf. section 6). Often an offset is applied to the virtual hand to avoid occlusion problems. This is not appropriate for the realistic interaction we focus at, since this would falsify the results of the functional validation. With this study we are investigating the influence of the virtual hand representation on the usability of our interaction metaphors. Of course, a virtual hand model cannot be avoided if an HMD is used because this type of display completely excludes the real world including the users' real hands.

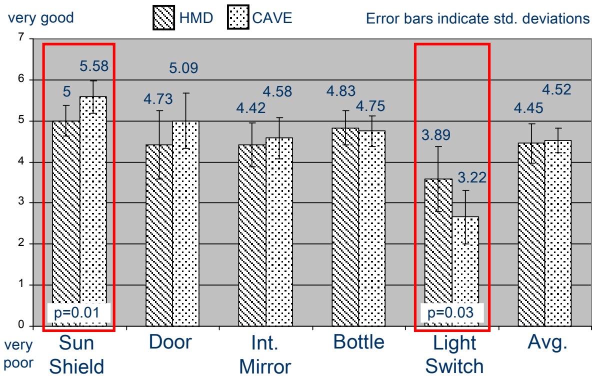

Figure 10: Per-object and subjective measures for HMD and CAVE.Framed columns represent at least weak significant differences.

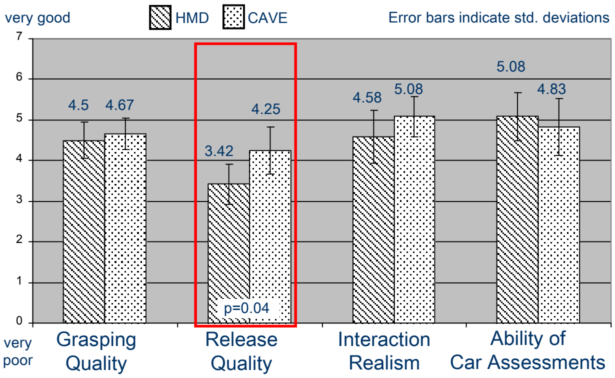

Figure 11: Questionaire results for HMD and CAVE. Framed Columns represent at least weak significant differences.

8.1 Methods

Participants and Apparatus

We performed this study together with the display evaluation study using the same 12 participants. Since this experiment only makes sense in a projection-based environment only the already described CAVE was used.

Design and Procedure

For this study we re-used the design of the display influence experiment. Again we calibrated the finger tracking system to the individual users' hands before each test. After a short familiarization procedure half of the group started the test with a virtual hand representation; the other half started without a virtual hand. The test under the first condition was directly followed by the other condition without removing the finger tracking system in between. After both trials the subjects were asked for their preference and any further comments. We did not use any proxies to improve grasp recognition during our tests to have a broad range of difficulties in our test to discover the limits of each condition. As a consequence the tiny light switch and the thin door handle were hard to grasp for some users.

8.2 Results and Discussion

User Preference and User Observation

The subjective results show a clear preference for bare hand interaction. Only one third of the subjects preferred the presence of a virtual hand representation.

The main advantage of the absence of a virtual hand was that interaction was perceived as being more direct and natural. The participants mentioned the direct relation between the users' real hand movements and the virtual object reaction as the main reason for rejecting the virtual hand representation. The virtual hand representation was

Figure 12: Interaction with the steering wheel with (left) and without (right) a virtual hand representation

said to be more indirect similar to a mouse cursor. Some subjects claimed that the virtual hand was not perfectly matching the real hand (Figure 12) and directs the users' attention to the hands instead of the objects. Since there is a distance between the real hands and the projection screen this can cause focus shifts that disturb the correct perception of the stereo images.

The advantage of the virtual hand representation could be clearly seen when the users interacted with difficult objects – the light switch and the door handle. For most of the users it was hard to grasp these objects without seeing the virtual hand. Five subjects were not able to interact with the light switch at all and three subjects had the same problem with the door in contrast to one and no subject respectively when the virtual hand was visible. The virtual hand representation provides more feedback on were the users' hands are located with respect to the objects and facilitate the interaction with challenging objects.

Task Completion Times

TCTs are comparable under both conditions for the objects that can be grasped without any problems (sun shield, mirror, bottle – see Figure 13). The inter- and inner-subject variances are low and similar for both conditions. The sun shield task could be completed significantly faster without a virtual hand representation (t=2.8, p=0.02, df=11) but with a minor difference of half a second.

The problems some users had with the light switch and the door were directly reflected by the task performance, leading to significantly higher task completion times in the no-hand condition (t=-2.2, p=0.05, df=10 and t=-2.2, p=0.05, df=11 respectively). Variances are very high for these objects in both conditions. In seven cases subjects were not able to interact with one of these objects in the no-hand condition. For calculating the mean values of task performance in this case we used the highest TCT the participant achieved for this particular object in hand condition as the TCT for the no-hand condition.

Subjective Measures

Subjective measures reflect the users' preference for the no hand condition resulting in better judgments for easy-to-use objects. For the door and the light switch the grasp problems lead to better judgments for the hand

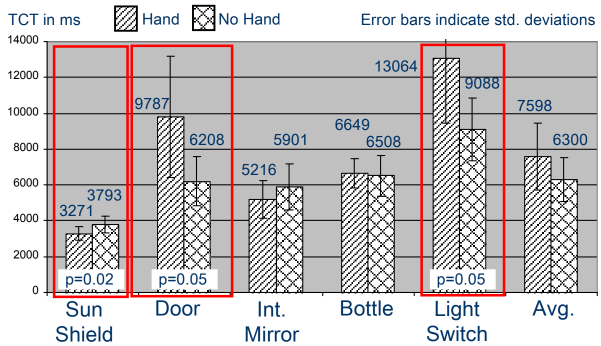

Figure 13: Per-object and average task completion times for the no hand vs. hand trial. Framed columns represent at least weak significant differences.

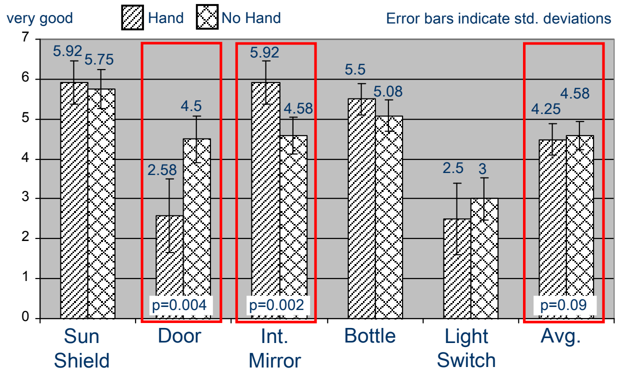

Figure 14: Per-object and average subjective measures for the no hand vs. hand trial. Framed columns represent at least weak significant differences.

condition (Figure 14). Concerning subjective measures the T-test showed a significant preference for the hand condition in the door task (t=3.6, p=0.004, df=11) and a significant better value for the mirror task (t=-4, p=0.002, df=11) when it was performed without a visible hand.

In general it can be stated that users do not prefer having a virtual hand representation. If grasping is robust they do not need it to judge the location of their hands with respect to the virtual objects. In fact having no virtual hand leads to a higher perceived realism and naturalness of the interaction metaphor. In difficult cases a virtual hand can help users to understand why an object cannot be properly grasped. Of course, without a virtual hand, the projection-based virtual environments have to have precisely calibrated screen setups and tracking systems.

9 Conclusions and Future Work

We introduced our approach towards Natural Interaction in automotive virtual environments. Our combination of robust grasping heuristics and pseudo-physical object behavior imitates physical processes found in the real world and enables the users to directly interact with freely moveable as well as constrained objects in a realistic and plausible manner. We analyzed the objects that can be found in our car interior scenarios and support a variety of typical constraints for these objects. One-handed interaction is possible with our approach as well as multi-hand and multi-user interaction. To increase the robustness of grasping, we introduced the concept of Normal Proxies, which extend the original object surfaces with appropriate normals for the grasping heuristics.

We studied our interaction metaphors in a CAVE and HMD environment and show that user performance and – surprisingly – preference were similar for both displays. A tendency to higher task performance was seen for the HMD while a slight subjective preference was reported for the CAVE. There is also a general preference in the CAVE for working without a virtual hand representation if the manipulation does not require a precise judgment of the hand position in relation to the virtual object. Overall, we found that users who have some experience with stereoscopic technology became quickly proficient with our interaction metaphors. Automotive experts could easily validate various functional aspects of a car, such as the accessibility and usability of car interfaces, visibility aspects and object clearances. However, the limitation is that task completion times are two to four times longer than in a real car.

The pseudo-physical reaction of objects could be further extended to provide several pseudo-haptic effects [7]. Different acceleration and inertia parameters could be applied to object reaction such that the object immediately follows the user's hand motion to simulate lightweight objects while heavy objects would exhibit more inertia. It is our belief that the range of pseudo-physical behavior for automotive applications is fairly exhausted with these extensions. Further physical effects, such as realistic object-object interaction, the influence of gravity and inertia or deformable hand models for more realistic hand behavior, can only be provided by physical simulations, which are still not robust enough for these challenging tasks. However, the gaming market as a driving force and further increases in readily available computational resources will enable real physical interaction metaphors in the not so distant future – even for our complex virtual scenarios.

REFERENCES

[1] D. Baraff, "Rigid Body Simulation", Course Notes of ACM Siggraph'97, Los Angeles (CA), USA, 1997.

[2] C.W. Borst, A.P. Indugula, "Realistic Virtual Grasping", Proceedings of IEEE Virtual Reality 2005, pp. 91-98, 2005.

[3] D.A. Bowman, E. Kruijff, J.L. LaViola, Jr., I. Poupyrev, "3D User Interfaces, Theory and Practice", Addison-Wesley, Boston, USA, 2005.

[4] B. Froehlich, H. Tramberend, A. Beers, M. Agrawala, D. Baraff, "Physically-based Manipulation on the Responsive Workbench", Proceedings of IEEE Virtual Reality'00, New Brunswick (NJ), USA, pp. 5-11, 2000.

[5] D. Holz, S. Ullrich, M. Wolter, T. Kuhlen, "Multi-Contact Grasp Interaction for Virtual Environments", Journal on Virtual Reality and Broadcasting JVRB, Vol.5 No.7, Köln, Germany, 2008.

[6] S.B. Kang, K. Ikeuchi, "A Framework for Recognizing Grasps", Techn. Rep. CMU-RI-TR-91-24, Carnegie Mellon University, Pittsburgh (PA), USA, 1991.

[7] A. Lecuyer, S. Coquillart, A. Kheddar, P. Richard, P. Coiffet, "Pseudo-haptic feedback: can isometric input devices simulate forcefeedback?" Proceedings of IEEE Virtual Reality VR'00, New Brunswick (NJ), USA, pp. 83-90, 2000.

[8] M. Moehring, B. Froehlich, "Pseudo-Physical Interaction with a Virtual Car Interior in Immersive Environments", Proceedings of IPT'05, Aalborg, Denmark, pp. 181-191, 2005.

[9] J.J. Preece, Y. Rogers, H. Sharp, "Interaction Design: beyond human computer interaction", Wiley & Sons, New York (NY), 2002.

[10] K. Shoemake, "Animating rotation with quaternion curves", Proceedings of 12th Int'l Conference on Computer Graphics and Interactive Techniques, New York (NY), USA, pp. 245-254, 1985.

[11] G. Zachmann, A. Rettig, "Natural and Robust Interaction in Virtual Assembly Simulation", Proceedings of 8th ISPE Int'l Conference on Cuncurrent Engineering Research and Applications (ISPE/CE2001), 2001.

[12] H. Salzmann, B. Froehlich, "The Two-User Seating Buck: Enabling Face-to-Face Discussions of Novel Car Interface Concepts", Proceedings of the IEEE Virtual Reality 2008, pp. 75–82, 2008.

[13] B. Ries, V. Interrante, M. Kaeding, L. Anderson, "The effect of self-embodiment on distance perception in immersive virtual environments", Proceedings of the ACM symposium on Virtual reality software and technology (VRST) 2008, pp. 167-170, 2008.

[14] S. Zhai "Human Performance in Six Degree of Freedom Input Control", PhD Thesis, University of Toronto, 1995.

[15] J. Kjeldskov, "Combining Interaction Techniques and Display Types for Virtual Reality", Proceedings of OzCHI 2001, Edith Cowan University Press, Churchland, AUS, pp. 77-83, 2001.

[16] D.A. Bowman, A. Datey, U. Farooq, Y.S. Ryu, O. Vasnaik, "Empirical Comparisons of Virtual Environment Displays", Tech. rep., Virginia Tech Dept. of Computer Science, TR-01-19.

[17] W. Qi, R.M. Taylor II, C.G. Haley, J.-B. Martens, "A Comparison of Immersive HMD, Fish Tank VR and Fish Tank with Haptics Display for Volume Visualization", Proceedings of 3rd Symposium on Applied Perception in Graphics and Visualization, Boston (MA), USA, pp.51-58, 2006.

[18] C. Demiralp, C.D. Jackson, D.B. Karelitz, S. Zhang, D.H. Laidlaw, "CAVE and Fishtank Virtual-Reality Displays: A Qualitative and Quantitative Comparison", in IEEE Transactions on Visualization and Computer Graphics (TVCG), Vol.12, No.3, pp. 323-330, 2006.

[19] M. Moehring, B. Froehlich "Enabling Functional Validation of Virtual Cars Through Natural Interaction Metaphors", Proceedings of IEEE Virtual Reality 2010, pp. 27-34, 2010.

[20] E. Burns, S. Razzaque, A. Panter, M.C. Whitton, M.R. McCallus, "The Hand Is More Easily Fooled than the Eye: Users Are More Sensitive to Visual Interpenetration than to Visual-Proprioceptive Discrepancy", Presence: Teleoperators and Virtual Environments Vol.. 15, No. 1, pp. 1-15, 2006.

Mathias Moehring Received his Diploma Degree in Media Informatics from Bauhaus-Universität Weimar, Germany in 2005. Since then he has been a Ph.D student at Bauhaus-Universität and working for the Virtual Technologies team of Volkswagen Group Research. His main research topic is VR interaction in industrial applications. His research interests also include VR displays and computer graphics, tracking systems and Augmented Reality.

Bernd Froehlich is a full professor with the Media Faculty at Bauhaus-Universität Weimar, Germany. From 1997 to 2001 he held a position as senior scientist at the German National Research Center for Information Technology (GMD), where he was involved in scientific visualization research. From 1995 to 1997 he worked as a Research Associate with the Computer Graphics group at Stanford University. Bernd received his M.S. and Ph.D. degrees in Computer Science from the Technical University of Braunschweig in 1988 and 1992, respectively. He has served as a program co-chair for the IEEE VR, 3DUI, IPT/EGVE and VRST conferences. He is also a co-initiator of the 3DUI symposium series and he is the 2008 Virtual Reality Technical Achievement Award winner. His research interests include real-time rendering, visualization, 2D and 3D input devices, 3D interaction techniques, display technology and support for collaboration in co-located and distributed virtual environments.