

Finger-Based Manipulation in Immersive Spaces and the Real World

Emmanuelle Chapoulie *
Inria

Theophanis Tsandilas
Inria, Univ Paris-Sud

Lora Oehlberg
Inria, Univ Paris-Sud

Wendy Mackay
Inria, Univ Paris-Sud

George Drettakis
Inria

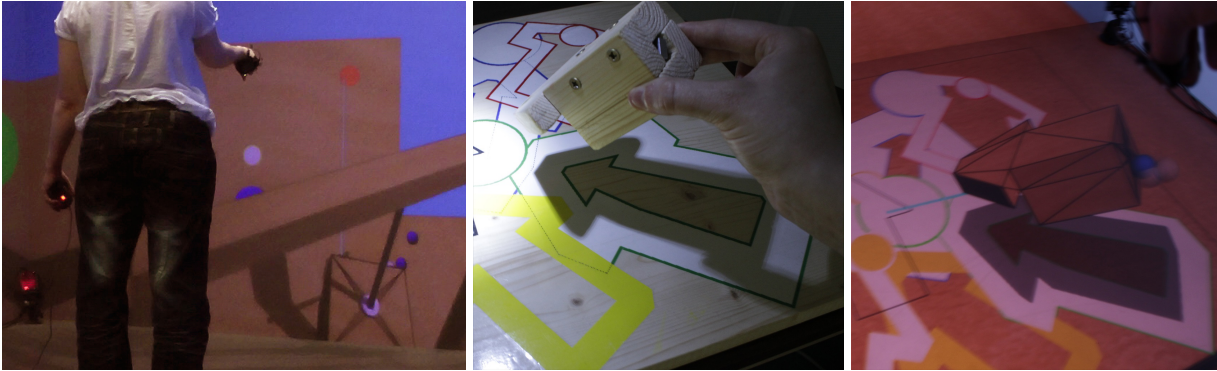


Figure 1: A user in our immersive environment (left). Completing a 6 DoF manipulation task in real (center) and virtual (right) settings.

ABSTRACT

Immersive environments that approximate natural interaction with physical 3D objects are designed to increase the user’s sense of presence and improve performance by allowing users to transfer existing skills and expertise from real to virtual environments. However, limitations of current Virtual Reality technologies, e.g., low-fidelity real-time physics simulations and tracking problems, make it difficult to ascertain the full potential of finger-based 3D manipulation techniques.

This paper decomposes 3D object manipulation into the component movements, taking into account both physical constraints and mechanics. We fabricate five physical devices that simulate these movements in a measurable way under experimental conditions. We then implement the devices in an immersive environment and conduct an experiment to evaluate direct finger-based against ray-based object manipulation. The key contribution of this work is the careful design and creation of physical and virtual devices to study physics-based 3D object manipulation in a rigorous manner in both real and virtual setups.

Index Terms: Immersive Cube-like Displays; Finger-based manipulation; Real/virtual world comparison.

1 INTRODUCTION

One of the key goals of an immersive cube, or CAVE, is to make the user’s experience as realistic as possible by simulating how the physical world responds to the user’s actions. The CAVE’s 3D graphics are enhanced with 3D manipulation techniques that simulate interaction with the physical world, thereby increasing the user’s sense of presence [23] and taking advantage of training and expertise gained in the physical world [4]. We focus on two approaches for manipulation in Virtual reality (VR), each with different limitations. The first lets users manipulate virtual objects directly with the fingers, which feels more “natural” and corresponds

to how we interact with physical objects in the real world. However, current VR systems have difficulty in accurately tracking the motion of hands and fingers in real time, and do not provide realistic haptic feedback that reflects relevant physical characteristics, such as object resistance and weight. In addition, some objects are difficult to reach, either because they are occluded by the hands or are out of reach.

Other approaches [2] use ray-casting, in which a wand or other device projects a ray onto the object of interest. Although less natural than using the hands, ray-based techniques make tracking easier by reducing the number of points tracked. They also avoid occlusion problems by allowing users to interact at a distance.

We are interested in understanding the trade-offs of each approach, given current technology limitations, and how they each compare to the manipulation of real objects in the physical world. Finding a representative task to study this problem is not trivial, since object manipulation is subject to a wide variety of factors. Scenes become increasingly complex as additional objects and surfaces, each with distinct roles and behaviors, are added and interconnected. We tackled the problem by decomposing 3D object manipulation into its component movements, subject to different physical constraints. This allows us to examine the following questions:

1. Can we build equivalent physical and virtual devices that decompose physics-based 3D-object manipulation into movements with varying degrees of freedom (DoF)?
2. Can we use these devices to study interaction in virtual environments and contrast it with physical interaction?

This paper proposes a framework for analyzing 3D object manipulation by first decomposing 3D movement into two movement types (rotation and translation) with one or multiple degrees of freedom (1D and 3D) tasks, and then building custom-built physical devices that permit the identical forms of virtual 3D object manipulation, subject to the identical constraints, as in finger-based and wand-based virtual environments. The set of physical devices provides a baseline of comparison for the two virtual techniques, and makes it possible to design experiments that control for factors such as target size, distance and feedback, with common measurements across all tasks. This provides a first example of how to systematically study 3D object manipulation in immersive environments

*e-mail: emmanuelle.chapoulie@inria.fr

that simulate physics and involve physical constraints. We demonstrate the approach by conducting a controlled experiment in which 16 participants performed a set of object manipulation tasks with the finger-based and wand-based virtual techniques, as well as the physical devices. We present the results of the experiment and discuss the strengths, limitations and trade-offs between finger-based and ray-based object manipulation in immersive environments. Finally, we discuss the advantages and limitations of this approach to studying interaction in physical and virtual environments.

2 RELATED WORK

In their thorough review of natural gestures for virtual reality, Bowman et al. [4] classify user interfaces into three main categories: *traditional 2D interfaces* such as classic mice and keyboards, *natural interfaces* where the user behaves as in daily life, and *hyper-natural interfaces* which extend the human capacities or provide guidance. *Hyper-natural interfaces* enhance interaction by making it more powerful, using a virtual or physical tool. On the contrary, *natural interfaces* provide interaction as close as possible to the real world through the use of the least invasive input devices possible. The authors underline the many positive features of *natural*, but also discuss the utility of *hyper-natural* interfaces. Indeed, this type of interaction may be more convenient than *natural interfaces* as it eases user actions, thus improving performance on task completion. For example, joystick-based interfaces are reported as being the most precise for 3D steering tasks because of rate-control. On the other hand, *natural interfaces* are mentioned as one of the important future directions for 3D user interfaces as they provide better spatial understanding when traveling, and better precision due to the scaling of movement. They also come easily for novice users as they are inherently intuitive, and ensure that training will transfer to the real world.

Considering finger-based interaction, Moehring and Froehlich [17, 18] develop a glove-based interface for functional aspect validation in automotive design. Users naturally interact with constrained objects with their phalanges. They can adjust an interior car mirror, press buttons and turn knobs in fully immersive settings. The system provides a robust and realistic heuristic-based interaction. The authors report that virtual reality allows the evaluation of generated models earlier in the design process, thus considerably reducing the number of design iterations and required hardware mockups. Also, Hilliges et al. [10] present an interactive system using a see-through display to provide the illusion of direct manipulation. Users interact within a “manipulation space” between the desktop and a half-silvered glass display. Their interface allows general purpose direct manipulation of virtual objects with bare hands, or with physical props. It also handles occlusions of virtual and real objects, providing a seamless mix.

Concerning the realism of the simulation, Sturman et al. [24] report that visual feedback alone cannot provide sufficient cues for gesture-based interaction, preventing users from experiencing direct presence in the virtual space. Similarly, in their system, Ullmann and Sauer [25] focus on the visual feedback quality to compensate as much as possible for the lack of haptic (force and touch) feedback. Within this perspective, Hirota and Hirose [12] introduce an algorithm for real-time haptic rendering with stable motion to simulate quite complex dexterous object manipulation. The user’s hand and fingers are represented by a set of interface points, and interaction force is computed for each of those. Friction can also be simulated. They observe from their evaluation study that force and tactile feedbacks are required for dexterous manipulations to reduce colliding volume and object slipping. Ortega et al. [20] propose a generalization to 6 DoF of the “God object” method for haptic interaction between rigid bodies. They provide high quality haptic interaction through continuous collision detection and constraint-based quasi-statics. Their method prevents interpenetration between ob-

jects and allows precise contact and sliding over surfaces. They report that haptics improve interaction because they allow to feel the detailed geometry of virtual objects. However, haptic devices are mechanical robots involving motors, such as the Phantom [16], the Pantograph [23] or the HapticMaster [26]. Thus, even if they greatly improve interaction with a virtual environment, they are restricted to relatively small workspaces such as desktop applications, being unsuitable for immersive cube-like displays.

The addition of physics simulation to direct manipulation provides natural feedback in the environment and thus truly allows for intuitive interaction with the objects in the scene. Such approaches for interface design have received much interest in recent years, often linked with tabletop systems. A major difficulty is how to handle objects controlled by the users hands (often called “God objects”) with respect to the simulation of the rest of the environment, i.e., correctly providing external forces from the hands. Borst and Indugula [3] use a spring model coupled with a commercially-available physics engine to simulate grasping and manipulation of objects of varying complexity; they avoid hand-object interpenetrations and provide force computation for haptic feedback. Grasping and interpenetration were also the focus of Prachyabrued and Borst [22]. Agarawala and Balakrishnan [1] propose a new and more physically realistic organizational structure for virtual desktops based on piling instead of filing. Wilson et al. [28] combine sensing surfaces and advanced game physics engines to provide highly realistic interaction with virtual objects. One remarkable early result is that by Froehlich et al. [8], which demonstrates the use of a fast physics solver and hand-based interaction, in the context of a workbench environment, for complex assembly tasks and multiple users and hands. The physics solver presented was one of the first providing sufficiently fast simulation to allow realistic interaction. More recent work has concentrated on developing appropriate soft models of fingers [14] and efficient solvers to avoid interpenetration of God objects and other objects in the scene [15], mainly in the context of automotive project review. Hilliges et al. [10] use physics particles and depth-aware optical flow to approximate 3D shapes of interacting rigid and non-rigid physical objects. A simpler approach was proposed by Holz et al. [13], where flexible grasping is simulated without complex physics, though allowing manipulation of multiple objects with multiple fingers. We follow a similar approach of simplified physics simulation.

When evaluating their urban planning system, Buchmann et al. [5] report user fatigue and frustration due to tracking problems. Similarly, O’Hagan et al. [19] warn about the importance of the chosen gesture set on the users’ fatigue when designing vocabulary-based interfaces. Zhai and Milgram [29] propose a new measure to quantify coordination in multiple degrees of freedom based on movement efficiency, and apply it to the evaluation of two devices: a free-moving position-control device and a desktop elastic rate-controlled hand controller. They conclude that more direct devices may lead to fatigue, coarseness of the control action (in terms of coordinated motion) and anatomical limitations of the human limb whereas they take shorter time to learn; on the contrary, less-direct tool-like devices may take more time to learn but may be more efficient. Finally, Graham and MacKenzie [9] compare user performance under a physical and a virtual setting. As opposed to our work, where we examine constrained 3D object manipulation, they focus on regular pointing tasks on a 2D plane. They find that the physical setting results in faster pointing. They attribute this difference to the limited visual feedback of their virtual setup.

3 OVERVIEW OF OUR APPROACH

As opposed to other virtual environments that normally extend the physical limits of reality, a main goal of immersive environments is to approximate interaction with our physical world. This implies that object manipulation for such environments should be studied

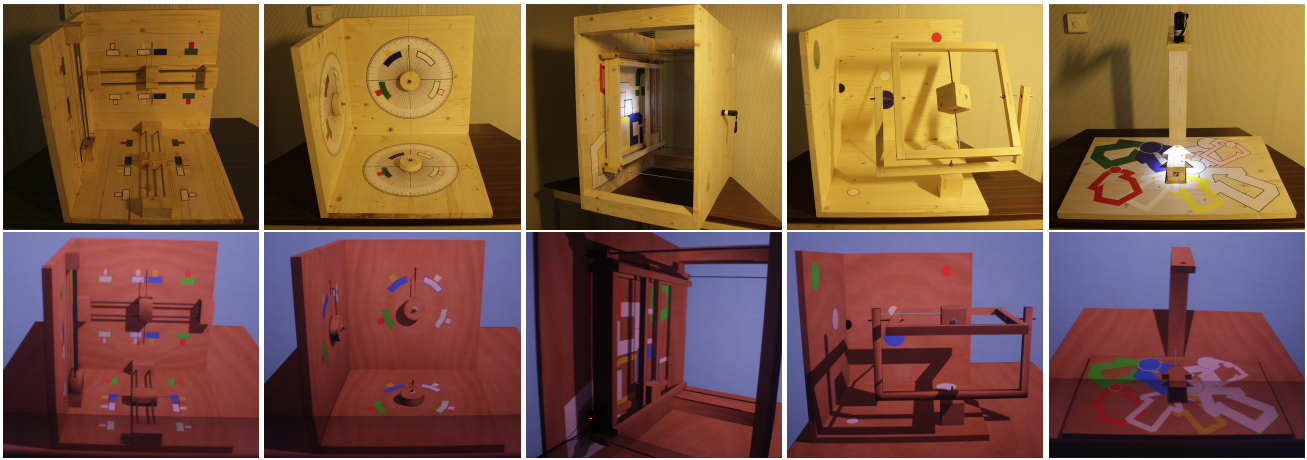


Figure 2: Devices used to decompose motion. From left to right: 1 DoF translations, 1DoF rotations, 3DoF translations, 3DoF rotations, free (6DoF) movements. The top line shows the physical devices, and the bottom line shows their virtual implementation.

with the laws and constraints of the physical world in mind. There are two directions to this goal. The common approach is to develop interactive VR technologies that simulate the physical world as close as possible. Unfortunately, current VR technologies are still far from providing a satisfying user experience when interaction comes into place. A second approach is to transfer the object-manipulation tasks from the virtual environment to the physical world and study them there. This approach has again limitations, as transferring a virtual 3D scene to the real world may require significant effort, time, and physical resources.

We consider both directions, trying to balance between the challenges they pose and their shortcomings. Instead of studying 3D object manipulation in complex real-world tasks, we decompose it into simpler constrained subtasks. These subtasks isolate translation and rotation components into a single or multiple DoF. More specifically, we study five constrained subtasks: (1) 1DoF translations, (2) 1DoF rotations, (3) 3DoF translations, (4) 3DoF rotations, and (5) 6DoF movements.

For each of these subtasks, we design and construct a physical and a virtual device that constrain 3D motion and can serve as apparatus in experiments that study 3D object manipulation under physical constraints. The five physical and virtual devices are shown in Fig. 2. The devices capture common experimental tasks that test movement performance: pointing and free or constrained 3D docking tasks. As we wanted to study such tasks within the context of real environments, our physical devices are subject to physical laws, e.g., friction and weight forces. Clearly, such phenomena produce confounding effects that can be hard to control in strict experimental settings. However, we consider this as a strength rather than a limitation of our approach, as it allows for studying user performance with natural tasks that obey the laws of physics.

4 INTERACTION IN THE IMMERSIVE ENVIRONMENT

Our virtual setup is a four-sided immersive space (three walls and a floor) using rear-projected black stereoscopic screens, also known as CAVE or “cube-like” immersive systems. Such setups have the advantage of providing full immersion over the entire field of view, while simultaneously allowing the user to perceive their body during interaction, improving the overall sense of presence. Our physics simulation uses the heuristic approach of Chapoulie et al. [6]. In addition to classic collisions, however, we also simulate the complex mechanisms of the five devices. We use the Bullet physics library to detect collisions, while we explicitly compute gravity and the mechanisms of the five devices based on simple mechanics.

We support two forms of interaction: (1) direct *finger-based* ma-

nipulation, and (2) ray-based manipulation with a *virtual wand*. Finger-based manipulation benefits from a representation of the fingertips as small colored spheres. We compensate for the lack of haptic feedback with visual feedback that shows when an object is grabbed. We highlight selected objects with a black and thick wire-frame visible from a wide range of viewpoints. The wand consists of a flystick that is extended with a virtual ray.

For tracking, we use an infrared optical system with eight cameras by ART. Finger-tracking devices are calibrated for each user. Although we use a high-end tracking system, the signal for the fingers is still sensitive to noise and interruptions. Most losses of signal are due to visibility problems that cannot be addressed: a user who walks and moves freely can often enter or be close to “blind regions” of trackers where the signal is deteriorated. This is primarily due to occlusions from the walls of the immersive cube as well as occlusions from the user. Other signal artifacts are subject to the reliability of finger markers. While the palm has many markers, each finger has only two, which results in unstable captured data.

Filtering the signal involved additional latency between the tracking data and the visual output. This can be perceived as elasticity between the users’ fingers and the object. A solution for similarity with real conditions could be to top the mobile objects of the physical devices with rubber parts to simulate the extra elasticity. However, we left this for future work.

5 FABRICATION PROBLEMS AND SOLUTIONS

The movement constraints of each physical device are to some extent subject to the fabrication tools and materials that were available. Building the devices, we had to make several compromises:

Robustness and Weight. Devices have to be robust to work under experimental settings that involve a large number of task repetitions with several users. At the same time, they have to be light enough to account for user fatigue but also to be close enough to the virtual devices where haptic feedback is absent. We chose to build the devices of pine and use stainless steel axes. However, due to the length of the axes in the 3D translation device, elasticity remained so we add frames to strengthen the mechanism.

Friction. Friction can also be a problem, so we tried to minimize its effect in our physical setup. We use needle-bearings for rotational and lubricated sliders for translational constraints.

Visual Feedback and Visibility. For movements with 3D translations, users must be able to easily control their placement, looking at a single position. To do so, we decided to use 2D projections of the movable object. We use a light source and shadow for

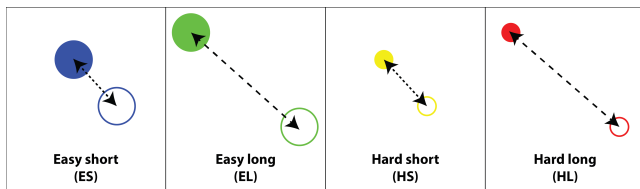


Figure 3: Color-coding of targets. Solid colors correspond to initial positions.

this (see Fig. 1) because the shadow position and its size are direct mappings of the 3D position of the object with respect to the source light. According to this solution, the user has to move the object so that its projected shadow fits into a specified target.

Despite our efforts to make physical and virtual as similar as possible, our solution has limitations, as user hands and the frames around the movable object also project shadows. Lighting conditions are thus important, so are the dimensions of the devices, as we must avoid shadow overlay while their parts move.

6 THE FIVE DEVICES

Our designs are based on simple but well-established mechanical systems that constrain the degree of freedom of movement. The devices simulate pointing and docking tasks and are equipped with visual instructions that indicate start and target task configurations. A target is a set of two colored areas: an area filled with *solid color* and a white area with an *outline color*. An elementary task consists of moving the device from a *solid color* to an *outline color*. The same device can be used for more than one task configurations.

We differentiate between tasks of a different difficulty level by coding targets with a different color. In our implementation, we use four colors (blue, green, yellow, red) to represent four difficulty levels (see Fig. 3), resulting from the combination of two target sizes (large vs. small) and two movement amplitudes (long vs. short).

6.1 1DoF Translations

This device is composed of three perpendicular panels, each presenting a linear slider oriented along a different direction. The slider mechanism is based on double-axis guidance from assembly machinery (see Fig. 2). The contact between wood and steel generates low friction, which is further reduced with lubrication.

The user faces the front panel and manipulates movable wooden parallelepipeds featuring stainless steel needles perpendicular to the translation axis. The targets are rectangular areas along the needle path. The user places the mobile parts so that the needles point to the inside of these areas (see Fig. 4).

6.2 1DoF Rotations

As the previous device, this device is composed of three panels, where each panel presents a wooden disk with a stainless steel needle (see Fig. 2). The disks act as household appliance knobs. Each knob rotates around a different axis. Friction is minimized by needle-bearings between the disks and their axes.

Again, the user faces the front panel to manipulate the knobs. Targets are circular areas along the path of the needle. The user can rotate a knob to make its needle point to a target (see Fig. 5).

6.3 3DoF Translations

This device is an extension of the 1D mechanism to 3D (see Fig. 2). A carter allows a frame to translate along the x direction through four parallel axes. This frame allows a smaller frame to move along the y direction through two parallel axes. Finally, that smaller frame permits a cube to translate along the z direction through two parallel axes. The user only manipulates the cube, which naturally triggers the movement of the other device components. Clearly, such translations along the three axes are not free, as movement is constrained

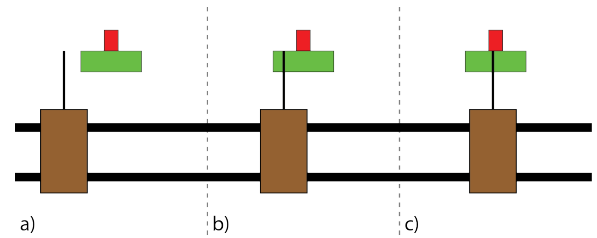


Figure 4: Placement check for 1D translations. The placement is: (a) incorrect, (b) correct for the green target only, and (c) correct for both the green and red targets.

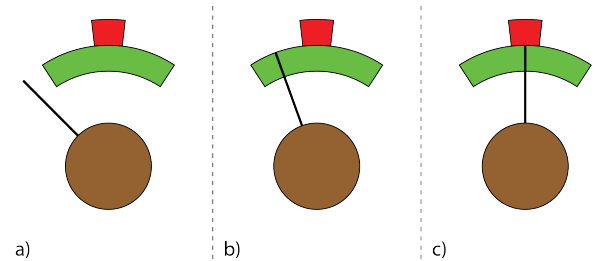


Figure 5: Placement check for 1D rotations. The placement is: (a) incorrect, (b) correct for the green target only, and (c) correct for both the green and red targets.

by the mechanical parts of the device. However, if we ignore weight and friction, constraints are theoretically identical for both the physical and the virtual devices.

Despite the robustness and rigidity of the stainless steel axes, the size of the physical device and thus the length of the axes implied some elasticity. Hence, we decided to strengthen the mechanism with extra wooden frames. Due to the increased weight to manipulate, we rotated the device to its side so that the user only needs to counterbalance for the weight of the graspable cube.

We use projected shadows for 3D positioning: the right panel holds a lamp and the object casts a shadow onto the left panel. When at a target position, the contour of the shadow is extended to form an outer contour and shrunk to form an inner contour. The combination of these contours generates a patch. To reach the target, the user has to place the object so that its projected shadow is entirely inside this patch, covering the inner contour, and not exceeding the outer contour (see Fig. 6).

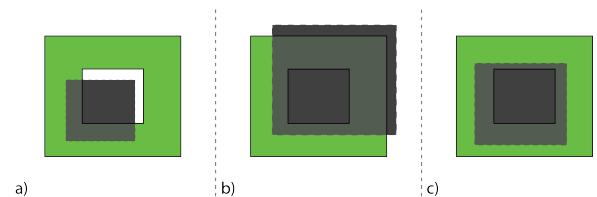


Figure 6: Shadow task, placement check. (a) The shadow does not completely cover the inner surface, the placement is not correct. (b) The shadow exceeds the outer contour, the placement is not correct. (c) The shadow contour is completely comprised between the inner and outer contours, the placement is correct.

6.4 3DoF Rotations

The 3D rotation device is a gyroscope (see Fig. 2). A central cube rotates around a y -oriented axis held by a square frame. This frame is held by a U-shaped frame and rotates around the x axis. This last frame is held by the carter and rotates around the z axis. Here again, needle-bearings are used to minimize friction. We know that such a mechanism is prone to gimbal lock; however, this does not interfere with the manipulation since the user can easily rotate the cube to escape from this configuration and restart the task.

| | Target Size | | | Movement Amplitude | | |
|-----------|--------------------------------|------|--------|--|--------|---------|
| | Description | Easy | Hard | Description | Short | Long |
| T1 | Patch length | 4 cm | 1.5 cm | Distance between patch centers | 5 cm | 24 cm |
| T2 | Patch angular length | 40° | 10° | Angular distance between patch centers | 72° | 255° |
| T3 | Patch thickness | 5 cm | 2 cm | Distance between 3D target positions | 13 cm | 31.4 cm |
| T4 | Dot radius | 4 cm | 2 cm | Angle between dot centers | 36.5° | 102.8° |
| T5 | Patch thickness and dot radius | 4 cm | 2 cm | Distance between 3D target positions | 9.7 cm | 27.8 cm |

Table 1: Target size and movement amplitude for each device.

As feedback for guiding the task, the cube features a red laser that replaces the needle of the 1DoF devices. The user has to point inside dot targets placed onto the carter panels.

6.5 6DoF Movements

The 6DoF movement task combines 3D rotation and translation. It requires no mechanics as the user freely moves a mobile object in space. Shadow targets are displayed on a plank featuring a bracket holding a light (see Fig. 2). We chose an anisotropic shape for the mobile object to disambiguate between orientations.

The body of the object holds a laser emitter. Targets combine a dot and a contour patch. To reach a target, the laser must point inside the dot while the projected shadow must be inside the patch.

7 EXPERIMENT

We conducted an experiment that compared finger-based to ray-based object manipulation of 3D objects with tasks of varying degrees of freedom. The two techniques were contrasted to natural object manipulation in a physical environment. The experiment provides a concrete example of how our framework can be used to study 3D interaction in immersive environments in comparison with interaction in the real world.

7.1 Participants

16 volunteers, 12 men and 4 women, 24 to 41 years old ($M = 28.7$ years, $SD = 4.9$ years) participated. Three participants had significant experience with VR environments while the rest had no or little experience. All were right-handed and had normal vision in terms of both color and stereo vision.

7.2 Techniques and Apparatus

The experiment tested two techniques for manipulating virtual objects: finger-based (*fingers_vr*) and ray-based manipulation with a wand (*wand_vr*). The two virtual techniques were contrasted to natural finger-manipulation of real objects (*fingers_real*). As a result, each participant was exposed to a total of three main conditions.

We used the virtual and physical devices that we presented in the previous sections. The hardware used for the virtual devices is a four-sided immersive projection system with Infitec stereo and ART tracking both for the head and fingers, and featuring a surround sound system.

To assess the effect of the tracking problem on the virtual techniques, we conducted a small noise study, where a user had to follow a guide along a path, keeping her hand steady. We measured the position of her hand, fingers and of the wand compared to the guide. We observed that the noise pattern seems to be similar for the wand and palm. However, finger-tracking introduced additional noise, whose order of magnitude is in centimeters. To overcome this problem, we apply a double filter, based on Kalman filter. We apply it to each finger. Details of the noise profiles are given as supplemental material.

7.3 Tasks

For each of the three main conditions, participants completed five types of tasks, which correspond to our five virtual and physical devices: 1DoF translations (T1), 1DoF rotations (T2), 3DoF constrained translations (T3), 3DoF constrained rotations (T4), and 6-DoF movements (T5). T1 and T2 had the form of regular 1D-pointing tasks. T3 and T4 were constrained, requiring participants to plan ahead and if necessary correct the path of the movement during a trial. More specifically, T3 was a 3D-pointing task and T4 was a constrained docking task. Finally, T5 was a regular free-movement docking task.

For all the five tasks, participants have to move a virtual or physical 3D object from a start to a target position. We control the difficulty of each task by varying the size of the target and the distance between the start and target positions. We organize trials as reciprocal movements where participants alternately move a 3D object between two positions, pressing a button each time they reach a target. For 1D tasks, the participant has to move forth and back three times, which results in six repetitive trials. As tasks T3 to T5 were much longer, we included a single forth and back movement, i.e., two repetitive trials. An error occurs if the object is not within the target position when the user presses the button. A trial continues until the user correctly places the object to the target and presses the button.

We only study movement time so our task design does not consider the time needed to plan the movement and grasp an object. Each movement type has a neutral position to ensure that all participants start with the same configuration. Participants have to grasp the object and move it to the indicated position to start a trial.

7.4 Design

We followed a within-participants design. The order of presentation of the two virtual conditions was counterbalanced among participants. However, they all tested the physical devices first, since these serve as baseline. Our pilot tests showed that people can understand the tasks faster if they are first exposed to the physical devices. We expected that learning effects would be minor for this condition.

Tasks were grouped into three main groups within each condition. Each participant first completed the 1D tasks (T1 and T2), continued with the two constrained 3D tasks (T3 and T4), and ended with the free 3D task (T5). The order of the task within the first two groups was counterbalanced among participants. This means that half of the participants tried the order (T1, T2, T3, T4, T5) while the other half tried the order (T2, T1, T4, T3, T5).

For each task, we varied the target size (E:easy vs. H:hard) and the amplitude of the movement (S:short vs. L:large). This results in four combinations of target size and movement amplitude (ES, EL, HS, HL) that correspond to four different difficulty levels. The exact values of these variables are task-specific and are summarized in Table 1. For T1 and T2, we also varied the axis of movement (x , y , and z). In total, each participant completed 108 unique trials. In addition to the main trials, participants completed enough practice trials that allowed them to familiarize with each task and condition.

| | $MT_{finger} - MT_r$ | $MT_{wand} - MT_r$ | $MT_{finger} - MT_{wand}$ |
|-----------|----------------------|--------------------|---------------------------|
| T1 | [0.44, 0.83]* | [0.34, 0.68]* | [0.01, 0.25]* |
| T2 | [0.79, 1.42]* | [0.60, 0.91]* | [0.14, 0.56]* |
| T3 | [-17.2, -4.76]* | [-13.5, -3.16]* | [-4.86, -0.36]* |
| T4 | [2.78, 5.76]* | [1.28, 3.53]* | [0.30, 3.42]* |
| T5 | [-1.82, 6.90] | [-1.72, 8.34] | [-5.46, 3.90] |

Table 2: 95% CIs of mean differences measured in seconds between techniques for tasks T1-T5. CIs have been adjusted for three pairwise comparisons. MT_{finger} , MT_{wand} , and MT_r is the mean time for *finger_vr*, *wand_vr*, and *finger_real*, respectively. Asterisks represent statistically significant differences ($p < .05$).

| | $E_{finger} - E_r$ | $E_{wand} - E_r$ | $E_{finger} - E_{wand}$ |
|-----------|--------------------|------------------|-------------------------|
| T1 | [4.3%, 8.9%]* | [4.8%, 11.6%]* | [-6.0%, 2.2%] |
| T2 | [6.0%, 14.1%]* | [3.5%, 9.3%]* | [0.1%, 7.3%]* |
| T3 | [-0.2%, 21.9%] | [0.0%, 21.9%]* | [-13.3%, 9.4%] |
| T4 | [17.2%, 30.5%]* | [4.7%, 21.1%]* | [1.6%, 20.3%]* |
| T5 | [13.3%, 36.7%]* | [4.7%, 24.2%]* | [-3.0%, 23.4%] |

Table 3: 95% CIs of mean differences of error rates. CIs have been adjusted for three pairwise comparisons. They have been computed by bootstrapping (bias-corrected & accelerated method with 10000 samples). E_{finger} , E_{wand} , and E_r is the mean error rate for *finger_vr*, *wand_vr*, and *finger_real*, respectively.

7.5 Procedure

Participants first passed a stereo-blindness test. They were then introduced to the physical devices and took a brief break after the completion of this condition. The two VR conditions started by a calibration process that provided the settings for the *fingers_vr* and *wand_vr* techniques. Calibration included two steps: (1) an eye-position calibration to create a profile that corrects the default projection in the immersive space, and (2) a finger-tracking calibration to create a hand model specific to the user. For the first step, we combined the protocol of Ponto et al. [21] with the calibration test of Chapoulie et al. [6]. Participants were instructed to complete the tasks as quickly and as accurately as possible.

After the end of the experiment, participants completed a 22-item cybersickness questionnaire [27] and a questionnaire to evaluate their experience with the two virtual techniques. They were asked to rate similarity with real-world manipulation, ease of use, fatigue, performance (precision and speed), and general appreciation on a Likert scale. Each session lasted approximately 2.5 hours.

7.6 Measures

We measure the *Time* it takes to move an object from the start to the target position. We also measure *Error* expressed as % error rate. For the physical setup, time and errors were estimated from video captures by making use of mouse clicks on the audio track. For video analysis, we used the Atlas.ti (www.atlasti.com) software.

7.7 Results

Figures 7 and 8 show our results for movement time and errors. Participants' answers did not reveal any cybersickness problem. Furthermore, we did not observe any strong correlation or clear pattern between VR experience and user performance for the two virtual techniques, so we omit any further discussion below.

7.7.1 Main Analysis

We conduct a separate analysis for each task. For time measurements, we take the median of trial repetitions. For T1 and T2, we conduct $3 \times 2 \times 2 \times 3$ RM ANOVAs, where *Technique*, *Size* of target, *Amplitude*, and *Axis* of movement are handled as repeated measures. For T3 - T5, we conduct $3 \times 2 \times 2$ RM ANOVAs as the

Axis variable is not relevant in this case. For pairwise comparisons (see Tables 2 and 3), we report the 95% confidence intervals (CIs) of mean differences, where we account for multiple comparisons by using Bonferroni's adjustment.

T1 - 1DoF Translations. There is a significant effect of *Technique* on *Error* ($\chi^2 = 25.29$, $p < 0.001$) and *Time* ($F_{2,30} = 61.32$, $p < 0.001$). Table 3 shows that *fingers_real* produced less errors. Table 2 also shows that *wand_vr* is slower than *finger_real* but faster than *finger_vr*. As expected, main effects on *Time* are significant for *Size* ($F_{1,15} = 88.95$, $p < 0.001$) and *Amplitude* ($F_{1,15} = 135.11$, $p < 0.001$). Interestingly, the effect of *Axis* is also significant ($F_{2,30} = 6.78$, $p = 0.004$). Translations along the *x* axis are 62 ms (CI: [2, 122] ms) faster than translations along the *y* and 71 ms (CI: [10, 132] ms) faster than translations along the *z* axis.

T2 - 1DoF Rotations. There is a significant effect of *Technique* on *Error* ($\chi^2 = 24.03$, $p < 0.001$) and *Time* ($F_{2,30} = 83.31$, $p < 0.001$). For both *Time* and *Error*, *fingers_real* has the best performance, followed by *wand_vr*. Main effects on *Time* are again significant for *Size* ($F_{1,15} = 97.53$, $p < 0.001$) and *Amplitude* ($F_{1,15} = 386.96$, $p < 0.001$). Again, the effect of *Axis* is significant ($F_{2,30} = 33.04$, $p < 0.001$). Now, rotations on the *y* axis are 206 ms (CI: [109, 304] ms) faster than rotations on the *x* axis and 230 ms (CI: [171, 290] ms) faster than rotations on the *z* axis.

T3 - 3DoF Translations. There is a significant effect of *Technique* on *Error* ($\chi^2 = 7.73$, $p = 0.021$) and *Time* ($F_{1,2,17.7} = 20.25$, $p < 0.001$). Although *fingers_real* produced less errors, the time performance of the physical device is particularly low. Another interesting finding is that *fingers_vr* is significantly faster than *wand_vr* for this task. Finally, main effects on *Time* are again significant for both *Size* ($F_{1,15} = 42.46$, $p < 0.001$) and *Amplitude* ($F_{1,15} = 37.60$, $p < 0.001$).

T4 - 3DoF Rotations. There is a significant effect of *Technique* on *Error* ($\chi^2 = 19.00$, $p < 0.001$) and *Time* ($F_{2,30} = 33.75$, $p < 0.001$). For both measures, the physical device has again the best performance, while *fingers_real* has the worst. The main effect on *Time* was not found to be significant for *Size* ($F_{1,15} = 4.45$, $p = 0.052$) and *Amplitude* ($F_{1,15} = 0.19$, $p = 0.67$). The mechanical constraints of the device had a dominant role in this task. We suspect that these constraints increased variance, hiding the effect of our two independent variables of task difficulty.

T5 - Free 6DoF Movements. *Technique* had a significant effect on *Error* ($\chi^2 = 11.38$, $p = 0.003$) with *fingers_real* producing significantly less errors. However, its effect on *Time* is not found to be significant ($F_{2,30} = 1.98$, $p = 0.16$). However, there is a significant interaction effect *Technique* \times *Distance* ($F_{2,30} = 6.10$, $p = 0.006$), explained by the fact that the performance of *wand_vr* deteriorates faster under long amplitudes. Finally, main effects on *Time* are significant for both *Size* ($F_{1,15} = 52.04$, $p < 0.001$) and *Amplitude* ($F_{1,15} = 31.51$, $p < 0.001$).

7.7.2 Performance Models

As 1D tasks can be viewed as target acquisition tasks, the difficulty *ID* of the tasks can be expressed by using Fitts' formulation [7]:

$$ID = \log_2\left(1 + \frac{A}{W}\right) \quad (1)$$

where *W* is the width of the target and *A* is the amplitude of the movement. For rotational movements (T2), we consider angular widths and amplitudes. Table 4 presents the results of the linear regressions that estimate the model parameters for each technique. We observe that movement time fits well to the model with the exception of *fingers_real* for T2.

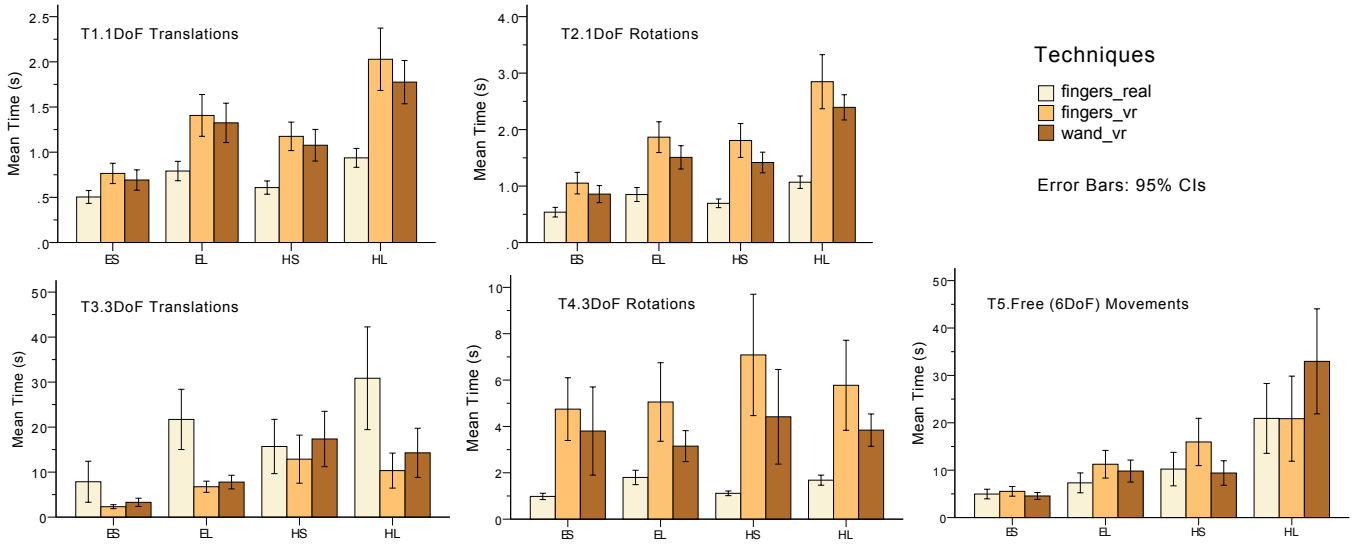


Figure 7: Results for movement time. Results are grouped by task and difficulty level.

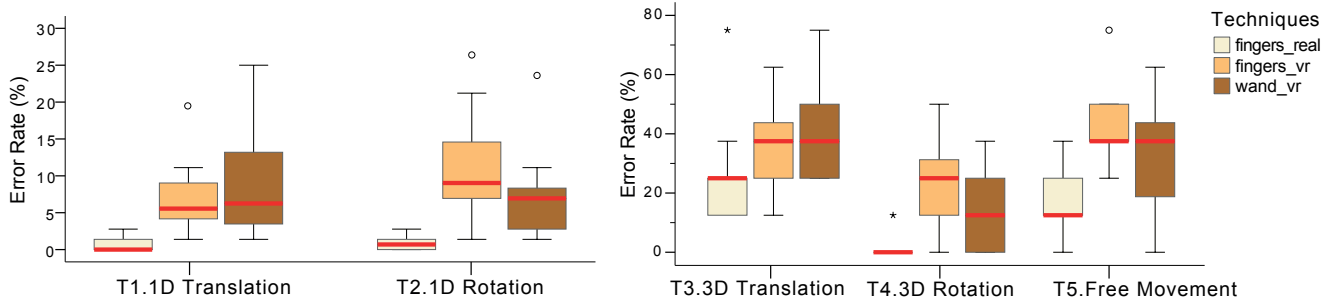


Figure 8: Box plots summarizing distributions of error rates for 1D (left) and 3D tasks (right). Red lines represent medians. Circles and asterisks represent outliers and extremes.

| Task | Technique | a (ms) | b (ms) | R^2 |
|------|------------------|-------------|------------|-------|
| T1 | $fingers_{real}$ | [177, 583] | [82, 281] | 0.968 |
| | $fingers_{vr}$ | [262, 587] | [425, 585] | 0.997 |
| | $wand_{vr}$ | [201, 655] | [322, 545] | 0.993 |
| T2 | $fingers_{real}$ | [-176, 932] | [-44, 423] | 0.859 |
| | $fingers_{vr}$ | [-171, 104] | [389, 927] | 0.982 |
| | $wand_{vr}$ | [-354, 992] | [281, 849] | 0.973 |

Table 4: Estimation of Fitts’s model parameters for 1D tasks. We show 95% CIs of the mean movement time MT , where $MT = a + b \cdot ID$.

7.8 Discussion

Our virtual conditions generated high error rates, especially for the 3D movement tasks where median rates ranged from approximately 10% to 40%. These rates are considerably higher than the ones of the physical devices, suggesting that errors were mostly due to the specific limitations of the virtual environment rather than the nature of the actual movements that we studied. For four (T1, T2, T4, and T5) out of the five tasks, the physical devices outperform the virtual ones in terms of errors and time, supporting our argument that physical devices can serve as baseline conditions for studying 3D object manipulation. Differences between the physical and virtual devices become more apparent at higher difficulty levels.

The device for 3DoF translations proved to be problematic. The inherent complexity of the physical device can explain this problem, as friction and weight start becoming a problem. Consider that the virtual devices did not simulate friction and weight forces.

Dealing with such problems is a challenge for future research. Which direction to follow to close the gap between virtual and physical devices (e.g., eliminate forces from the physical devices or add forces to the virtual ones?) is not straightforward as researchers have to balance between limitations of immersive environments and the constraints of physical devices and fabrication tools.

The $wand_{vr}$ technique generally performed better than $fingers_{vr}$, especially for rotational movements. However, in the case of 3DoF translations and free movements, results are less clear. Participants did not express any clear preference in favor of one technique or the other in terms of performance, fatigue, cybersickness, ease of use, and overall appreciation. On the other hand, they found that finger-based manipulation was closer to real object manipulation. Results are definitely not conclusive as the $fingers_{vr}$ technique was more prone to tracking problems and jumps due to the limited number of markers and their occasional occlusion from the user. Occlusion seems to be a more general problem for finger-based manipulation. Several participants reported that interacting with fingers caused more visibility problems than the wand, where object manipulation takes place at a distance. Visibility was more crucial in rotations, which can explain the poor performance of $fingers_{vr}$ for these tasks. In contrast, occlusion was not a problem in 3D translations due to targets being away from objects.

8 CONCLUSIONS AND FUTURE WORK

We introduced a framework for the analysis of 3D object manipulation through five physical and virtual devices that decompose movement into varying degrees of freedom: 1DoF translations, 1DoF

rotations, 3DoF translations, 3DoF rotations, and 6DoF free movements. We learned important lessons about the compromises required to design devices that (a) are reproducible in both real and virtual settings, and (b) can be used in experiments to measure user performance. The robustness of construction materials, weight, friction as well as visibility problems caused by moving parts are all factors with potential problems that researchers need to carefully consider when designing physical devices.

Creating a physical device for experimental testing poses additional challenges with no easy workarounds. How to control for task difficulty? How to show incremental feedback about the progress of a 3D task? How to make accurate measurements? We engineered and tested several solutions to these problems. For example, we used physical lighting and projected shadows to support visual feedback for 3D movements. These solutions are certainly limited in scope but present a first example of how we can create interactive physical and virtual task configurations that are measurable and comparable.

We used our devices in an experiment to investigate our second research question. Our results provide insights about finger-based and ray-based object manipulation for various types of movement constraints. Our proof of concept demonstrates the feasibility of our approach. As Graham and MacKenzie [9], we found that overall, the physical setups result in better user performance. For 3D tasks, however, the physical constraints of real-world devices make things more complex. This part of work needs further investigation.

Creating a virtual system that approximates natural interaction with physical objects remains a challenge. Finger tracking in cube-like immersive displays is problematic mainly due to "blind regions" or occlusions. Magnetic motion-tracking technologies avoid occlusion problems but are cumbersome and have a limited effective range. This situation won't get better any time soon, also because users have to move in a dark environment where vision-based systems cannot work. However, we believe that studying immersive environments in parallel with real ones can help us better understand their shortcomings and their future potential. It can also help us compare interaction techniques both in conjunction with and in separation from the limitations of current technologies.

We only examined finger-based and ray-based object manipulation in this work. An alternative solution is to use passive props [11], which alleviate the lack of haptic feedback. Yet, how to design generic props that effectively deal with different grasps and movement constraints is not a simple matter. Comparing props with real-world manipulation is an interesting future direction.

ACKNOWLEDGEMENTS

The authors wish to thank all the pilots and participants of the study. This work was supported in part by the Regional Council of Provence Alpes-Côte d'Azur and the EU project VERVE (<http://www.verveconsortium.eu>).

REFERENCES

- [1] A. Agarwala and R. Balakrishnan. Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. *CHI '06*, pages 1283–1292, 2006.
- [2] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers and Graphics*, 37(3):121–136, 2013.
- [3] C. Borst and A. Indugula. Realistic virtual grasping. In *IEEE VR 2005*, pages 91–98, 2005.
- [4] D. Bowman, R. McMahan, and E. Ragan. Questioning naturalism in 3d user interfaces. *Communications of the ACM*, 55(9):78–88, 2012.
- [5] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn. Fingertips: gesture based direct manipulation in augmented reality. In *Proc. GRAPHITE '04*, pages 212–221, 2004.
- [6] E. Chapoulie, M. Marchal, E. Dimara, M. Roussou, J.-C. Lombardo, and G. Drettakis. Evaluation of direct manipulation using finger

- tracking for complex tasks in an immersive cube. *Virtual Reality*, 18(3):203–217, 2014.
- [7] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.
- [8] B. Fröhlich, H. Tramberend, A. Beers, M. Agrawala, and D. Baraff. Physically-Based Manipulation on the Responsive Workbench. In *IEEE VR 2000*, 2000.
- [9] E. D. Graham and C. L. MacKenzie. Physical versus virtual pointing. In *CHI '96*, CHI '96, pages 292–299, New York, NY, USA, 1996. ACM.
- [10] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson. HoloDesk : Direct 3D Interactions with a Situated See-Through Display. In *CHI '12*, pages 2421–2430, 2012.
- [11] K. Hincley, R. Pausch, J. C. Goble, and N. F. Kassell. Passive real-world interface props for neurosurgical visualization. In *CHI '94*, CHI '94, pages 452–458, New York, NY, USA, 1994. ACM.
- [12] K. Hirota and M. Hirose. Dexterous object manipulation based on collision response. In *IEEE VR '03*, volume 2003, pages 232–239. IEEE Comput. Soc, 2003.
- [13] D. Holz, S. Ullrich, M. Wolter, and T. Kuhlen. Multi-Contact Grasp Interaction for Virtual Environments. *Journal of Virtual Reality and Broadcasting*, 5(7):1860–2037, 2008.
- [14] J. Jacobs and B. Froehlich. A soft hand model for physically-based manipulation of virtual objects. In *IEEE VR 2011*, IEEE, 2011.
- [15] J. Jacobs, M. Stengel, and B. Froehlich. A generalized God-object method for plausible finger-based interactions in virtual environments. In *3DUI'2012*, pages 43–51. Ieee, Mar. 2012.
- [16] T. H. Massie and J. K. Salisbury. The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems*, volume 55, pages 295–300. Chicago, IL, 1994.
- [17] M. Moehring and B. Froehlich. Enabling functional validation of virtual cars through natural interaction metaphors. In *Virtual Reality Conference (VR)*, 2010 IEEE, pages 27–34. IEEE, 2010.
- [18] M. Moehring and B. Froehlich. Natural interaction metaphors for functional validations of virtual car models. *IEEE TVCG*, 17(9):1195–1208, 2011.
- [19] R. O'Hagan, A. Zelinsky, and S. Rougeaux. Visual gesture interfaces for virtual environments. *Interacting with Computers*, 14(3):231 – 250, 2002.
- [20] M. Ortega, S. Redon, and S. Coquillart. A six degree-of-freedom god-object method for haptic display of rigid bodies with surface properties. *IEEE TVCG*, 13(3):458–469, 2007.
- [21] K. Ponto, M. Gleicher, R. G. Radwin, and H. J. Shin. Perceptual calibration for immersive display environments. *Visualization and Computer Graphics*, *IEEE Transactions on*, 19(4):691–700, 2013.
- [22] M. Prachyabrued and C. Borst. Visual interpenetration tradeoffs in whole-hand virtual grasping. In *3DUI*, pages 39–42. IEEE, 2012.
- [23] C. Ramstein and V. Hayward. The pantograph: a large workspace haptic device for multimodal human computer interaction. In *Conference companion on Human factors in computing systems*, pages 57–58. ACM, 1994.
- [24] D. J. Sturman, D. Zeltzer, and S. Pieper. Hands-on interaction with virtual environments. *UIST '89*, pages 19–24, 1989.
- [25] T. Ullmann and J. Sauer. Intuitive Virtual Grasping for non Haptic Environments. In *Pacific Graphics '00*, pages 373–381, 2000.
- [26] R. Q. Van der Linde, P. Lammertse, E. Frederiksen, and B. Ruiters. The hapticmaster, a new high-performance haptic interface. In *Proc. Eurohaptics*, pages 1–5, 2002.
- [27] I. Viaud-Delmon, Y. P. Ivanenko, A. Berthoz, and R. Jouvent. Adaptation as a sensorial profile in trait anxiety: a study with virtual reality. *Journal of anxiety disorders*, 14(6):583–601, 2000.
- [28] A. D. Wilson, S. Izadi, O. Hilliges, A. Garcia-Mendoza, and D. Kirk. Bringing Physics to the Surface. In *ACM UIST '08*, pages 67–76, 2008.
- [29] S. Zhai and P. Milgram. Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In *CHI '98*, pages 320–327. ACM Press, 1998.