IPT-EGVE Symposium (2007)
B. Fröhlich, R. Blach, and R. van Liere (Editors)

A Comparison of Tracking- and Controller-Based Input for Complex Bimanual Interaction in Virtual Environments

André Kunert Alexander Kulik Anke Huckauf Bernd Fröhlich

Bauhaus-Universität Weimar

Abstract

We describe a user study comparing a two-handed controller-based input device to a two-handed tracking solution, both offering the control space of six degrees of freedom to each hand. For benchmarking the different input modalities we implemented a set of evaluation tasks requiring viewpoint navigation, selection and object manipulation in a maze-like virtual environment. The results of the study reveal similar overall performance for both input modalities for compound tasks. However significant differences with respect to the involved subtasks were found. Furthermore we can show that the integral attributes of a subtask do not necessarily need to be manipulated by a single hand. Instead, the simultaneously required degrees of freedom for operating integrally perceived subtasks may also be distributed to both hands for better control.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation]: User Interfaces: Input Devices and Strategies

1. Introduction

Interaction in three-dimensional environments is typically a compound task consisting of navigation, object selection and manipulation as well as system control actions. Users often need to alternate between these basic modalities. Bimanual interaction has the potential for fast and modeless task switching by assigning different tasks to each hand. The left and right hand tasks could be even executed in parallel by a skilled user, e.g. during navigation an object could be picked up and moved to a new location. In addition different parameters of a single task might be manipulated in parallel by the left and right hand for better control.

For stereoscopic virtual environments direct six degree of freedom (DOF) tracking-based interaction is often considered a good choice, since it enables the implementation of easy to understand techniques such as ray-based selection and object manipulation as well as hand gesture driven navigation. Handheld input devices such as common game controllers are an interesting alternative, since they leverage the fine motor skills of the human hand and finger system. In addition such input devices also work in environments where no tracking is available. Even for basic tasks or task sequences little is known about the relative performance and usability of these different input modalities.

Our goal was to get insights into the usability of tracker-based interaction techniques in comparison to those controlled through handheld input devices. We developed a testbed for setting up and measuring various parameters of a variety of task sequences consisting of navigation, object selection and manipulation parts. Within this framework, we implemented and continuously refined a set of one-handed and two-handed interaction techniques for such compound tasks, which were adapted for tracking-based and controller-based input respectively. An extended user study with expert users provides initial insights into the strengths and weaknesses of the different interaction approaches.

Several evaluation frameworks are proposed for the examination of the specific usability of input devices for 3D interaction (e. g. [Zha95], [BJH99]). However, these frameworks do not focus on compound tasks or task sequences. Huckauf et al. [HSK*05] suggested the use of an extended docking task consisting of navigation and object manipulation for evaluating a set of two-handed input devices. We generalize this idea to include object selection as well as different navigation strategies. System control has been excluded so far since such tasks can often be seen as special cases of object selection and manipulation e.g. selection of a menu item or manipulation of a slider.

For tracking-based input we use two handles each containing an electromagnetic 6-DOF sensor (Figure 1).

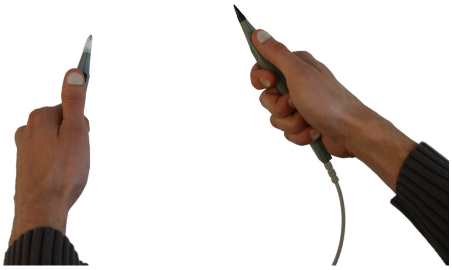


Figure 1: Tracking-based input

The handheld input device resembles a game controller but uses two small elastic 6-DOF sensors instead of the commonly found 2-DOF joysticks (Figure 2).

Figure 2: Controller-based input.

The results of our user study indicate that the tracking-based solution and the handheld controller-based input device perform on a similar overall level. However there are clear differences with respect to the task completion times for individual subtasks and in the temporal composition of compound tasks. The handheld input device performs better for manoeuvring tasks while the tracking solution performs better for object selection and manipulation. A closer analysis shows that the integral attributes of a basic subtask do not need to be manipulated by a single hand as described by Jacob et al. [JSPF94], but that the simultaneously required degrees of freedom may also be distributed to both hands provided that it is appropriately designed. These findings from our study help to understand the influence of the different input modalities on task performance and they provide initial hints for the design of appropriate interfaces for compound interaction tasks.

2. Related Work

Guiard [Gui87] classified manual activities into three categories: unimanual, bimanual symmetric and bimanual asymmetric. The bimanual asymmetric activities are by far the most common. They require a complex coordination between the hands. Guiard observed three high level principles for these tasks:

- The non-dominant hand adjusts a reference frame. The dominant hand acts upon this reference frame.
- The dominant hand produces fine-grained gestures. The non-dominant hand performs gross manipulations.
- The non-dominant hand initiates the action.

There are a variety of approaches for desktop interfaces trying to exploit the potential of bimanual interaction by assigning tasks to the left and right hands according to Guiard's findings. Buxton and Myers [BM86] equipped the users' dominant hand with a mouse for cursor-based selection and object manipulation and the non-dominant hand with a one-dimensional slider. Parallel bimanual input strategies were automatically adopted by novice users and led to better performance for a compound positioning/scaling task while for more complex interaction such as navigation and selection within a word document only few subjects adopted parallel input strategies - but those had the best results. Nevertheless, even for the compound navigation/selection task subjects performed faster with the bimanual technique. A proportional dependency between performance and simultaneity of two-handed input was found.

Jacob et al. [JSPF94] analysed the cognitive relation between the attributes of the application task and the control space offered by the input device. Based on Garner's work [Gar74] they showed that computer input devices providing simultaneously available DOF compatible with the structure of integral and separable attributes of a given task would perform better than those providing more or less independent input channels. On the other hand they also showed that for non-integral task attributes, such as position and colour, independently operable sensor devices are beneficial.

Considering two-handed input it seems that the theory of Jacob et al. is also applicable to the degrees of freedom simultaneously manipulated by two hands. It may even extend to any parallel input channel offered by a sensor network to the user, including speech input or sensors operated by the feet. In the experiments of Buxton and Myers [BM86] subjects immediately adopted parallel two-handed input strategies for a task requiring positioning and scaling of a displayed square - much the same task as chosen by Jacob et al. for the analysis of single input sensors. Their results showed that parallel two-handed input significantly improved performance.

Another experiment, conducted by Leganchuk et al. [LZB99], points at the impressing potential of designing the input modalities such that they allow for an effective chunking structure of the task. Chunking [Sim74] describes an organization model of the human memory alleging that information and/or cognitive skills related to each other by similarity or equivalence in certain attributes like context or spatial proximity are grouped into chunks. Lots of tasks in human computer interaction may be divided into separate subtasks or integrated into a single method chunk [CMN80]. Leganchuk et al. tested the performance of unimanual vs. bimanual methods for dragging out a selection mask exactly matching the targets' borders. Note that the manipulated attributes of the selection mask were the same as identified to be integrally perceived in the experiments of Jacob et al. [JSPF94]. The two-handed techniques allowed for parallel manipulation of position and size of the selection mask, which performed better than the one-handed approach. The simultaneity of input was shown to be beneficial for the task. A further aspect leading to shorter task completion times was the lower cognitive load due to the perception of the task as an integrated method chunk clasped by muscular tension to hold down and release the mouse button.

Balakrishnan and Kurtenbach [BK99] showed that bimanual 3D interaction is faster than one-handed, even if controlled attributes are not integrally related to each other. They analysed an interface design offering a 2-DOF mouse to each hand for camera control and object manipulation in 3D graphics interfaces. Their controller-multiplexing interface is based on Guiard's "kinematic chain model" of two-handed interaction and it showed significantly better performance compared to the common sequential one-handed approach. For a more complex compound task marginal benefits of two-handed input became only apparent after extensive practice.

Huckauf et al. [HSK*05] presented a comparison of two-handed input devices with each hand controlling six degrees of freedom. The dominant hand controlled the position and orientation of objects (manipulation) and the non-dominant hand controlled the viewpoint (navigation). They describe an extended docking task requiring both camera navigation and object manipulation. Following the findings of Buxton and Myers [BM86] two handed interaction that avoids frequent mode changes should increase user performance. Two-handed input should be especially beneficial if it allows the simultaneous operation of two subtasks. However, Huckauf et al. found no significant differences in their comparison to time sequential two-handed input even though there was a slight trend towards better performance for the simultaneous two handed input.

3. The Evaluation Scenario

The evaluation of input devices for specific subtasks such as object manipulation [Zha95] or navigation [Bow99] provides important information about the performance of input devices and interaction techniques. We suggest the individual evaluation of basic tasks followed by an evaluation of compound tasks consisting of common sequences of basic tasks to verify the transferability of the obtained results to such closer-to-reality situations.

3.1. Selection and Manipulation Tasks

While manipulation of three-dimensional objects often requires the control of six degrees of freedom (3D rotation and translation), selection using ray casting techniques may be operated with just two (pointing direction) or five (pointing direction plus ray origin) DOF. Another common way to select and manipulate objects is the virtual hand metaphor requiring three degrees of freedom for object selection. Based on the intuitiveness of the virtual hand metaphor we developed a novel 3D-cursor, offering additional orientation cues to improve the feedback for selected objects.

3.2. A novel 3D-Cursor

Our manipulation technique uses a ball-shaped 3D-cursor which is geometrically split up into eight segments. The cursor may be moved along the three spatial axes. If the cursor is moved inside the geometry of a displayed object, an animation process "explodes" the cursor, such that its separate parts surround the object's bounding sphere. Final grabbing via button click causes cursor adjustment to the object's orientation. The cursor now seems to clamp the selected object within its eight separated parts helping the users to recognize the orientation of the manipulated object. Using the cursor as a virtual manipulation tool, the object may then be adjusted in position and orientation as well. When released and moved out of the respective geometry, the cursor collapsed into its original shape and size. We implemented a visualization similar to the "silk cursor" proposed by Zhai et al. [ZBM94] using different states of transparency to handle occlusion.

3.3. Travel Tasks

Bowman et al. [BKVP05] identify three main travel tasks: exploration, search and manoeuvring. While search and exploration differ in the user's intention, both require similar movements to be controlled. Driving or flying metaphors, which allow for motion for- or backwards in an adjustable direction (heading), are mostly sufficient for this kind of travel tasks, conducted by users to reach distant locations. Manoeuvring on the other hand describes adjustments of the viewpoint on a smaller scale and is commonly required to examine virtual environments or objects from several perspectives. Movements perpendicular to the viewing direction (e.g. strafe, swing) are often applied in this context.

3.4. Hand-Eye Coordination as a Navigation Metaphor

Several bimanual interaction systems where, according to [Gui87], the results of input provided by the non-dominant hand serves as a reference for motion controlled by the dominant hand, were proved for high usability [HPP*97], [ZFS97]. All these systems considered only manipulation and viewpoint manoeuvring tasks, which are quite similar in terms of required motion control. Our research concerns interaction techniques to employ both hands for travel motivated by exploration and search. For applications focusing on manipulation tasks it makes perfect sense to distribute the high precision input required for object manipulation to the dominant hand. The control space of the dominant hand therefore contains three translational plus three rotational degrees of freedom (Figure 3).

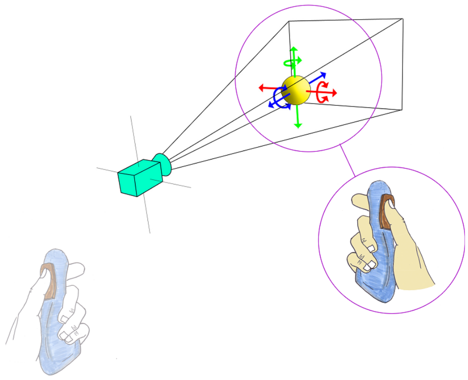


Figure 3: Dominant hand applied for manipulative operations.

The less demanding input for viewpoint control is applied to the non-dominant hand. At least two DOF (for-/backward, head) are required for distance travel. The more complex manoeuvring operations make use of two additional strafing directions and the pitch rotation, raising the necessary number of degrees of freedom to five (Figure 4).

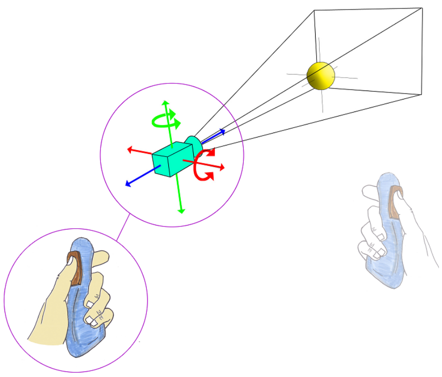


Figure 4: Non-dominant hand applied for viewpoint control.

Nevertheless, it is ineffective [Gui87] to explore simulated architecture or landscapes with an interface that constantly requires input from the left hand, while the right hand is excluded from interaction. In 3D shooter games two-handed input is successfully applied to search tasks within large virtual environments. The non-dominant hand defines one of four planar motion directions to move an egocentric avatar with constant velocity. The dominant hand controls a 2D cursor for the selection of focus and therefore to adjust the directions of motion, view and attack. This interaction concept, constraining viewpoint orientation to cursor translation, effectively minimizes the number of necessary degrees of freedom and therefore facilitates interaction. We assumed that our 3D-cursor could easily be employed in a similar way to define a focus point which serves for steering during travel. Adjusting the viewpoint orientation to a virtual tool for selection and manipulation somehow resembles human hand-eye coordination which describes the eye as being constrained to hand motion during manipulation tasks.

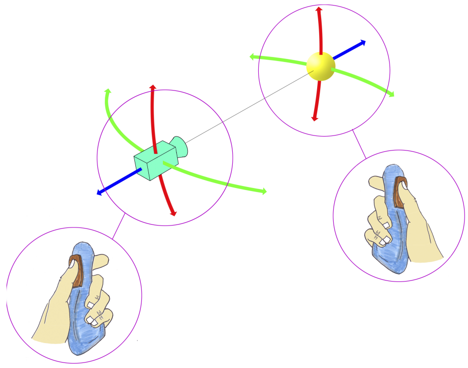


Figure 5: Bidirectional connection of camera and cursor.

We developed a technique to exploit simultaneous two-handed input for navigation tasks. The basic idea is to establish a geometric connection between the view and the cursor. Both are then forced to stay in line of sight towards each other (Figure 5). Hence sidewards translations of the cursor lead to viewpoint rotation of the attached camera, thus providing a pointing metaphor where the dominant hand input defines the steering direction for motion controlled by the non-dominant hand. Sidewards motion of the camera on the other hand results in encircling movement in relation to the cursor position. Thus all translations of viewpoint and cursor that are perpendicular to the line of sight between both, result in an encompassing movement around the respective other. For fluid switching between one-handed and two-handed interaction mode, we implemented a state model based on grabbing or releasing the corresponding physical controller device. Both hands' input may therefore be mapped to move viewpoint and cursor directly and individually during one-handed mode or in relation to each other during two-handed mode.

Objects are matter of manipulation related to the virtual environment. Thus it is not beneficial to circle them around the viewpoint. Encompassing viewpoint movements around the manipulated object on the other hand are quite helpful. With a selected object to be manipulated, the described technique therefore slightly changes its parameters. While object translations operated by the dominant hand are straight line movements, the identical translational input provided by the non-dominant hand causes the viewpoint to circle around the manipulated object. Following the idea that our objects have a certain mass, it remains at position during one-handed viewpoint control which assures that it will only be consciously moved and not by mistake.

3.5. Input Devices

We compared two types of input devices that strongly differ in their characteristics: a tracking-based system vs. a controller-based device.

3D tracking devices are commonly employed to generate motion input through large scale movements with the hand involving arm and shoulder. Alike the 2D computer mouse, 3D tracking devices belong to the group of isotonic sensors used for position control of virtual tools or objects. For our study we used two "Polhemus Fastrak" tracking devices (Figure 1), which represent the current state of the art among electromagnetic tracking devices.

Common video game controllers providing just two DOF to each hand are not comparable to tracked 6-DOF handles. Simon and Fröhlich introduced the Yoyo [SF03], a controller-based handheld device that consists of two "Spacemouse" force/torque sensors offering six degrees of freedom to each hand. This configuration of input sensors might compete with two tracked 6-DOF handles, but its design does not allow for parallel input of both hands, which was shown to be an important factor for the performance advantages of two-handed interaction [BM86]. Consequently the Square Bone, a similar device, providing the same input sensor configuration with a slight change of design to enable simultaneous use of both input sensors was developed [HSK*05]. For our study we added a new design to the Yoyo device family. We use smaller 6-DOF force/torque sensors (3D Connexion SpaceTraveller), which result in a device design resembling conventional game controllers in shape and size(Figure 2). We simply call the device "Bone", referencing the mentioned Square Bone.

The main differences between the two tested input systems are the following:

- While the elastic devices provide passive force feedback, the tracking devices are free floating (isotonic).
- The two input sensors of the Bone are physically attached to a single housing that is fixed between both hands, thus allowing only for fine motion input through the fingertips. The two separate handles of the tracking system, that may freely be moved throughout a large area, sense coarse and fine grain motion of both hands.
- The Bone consists of elastical sensor devices used for rate control of motion following [Zha95]. The isotonic tracking devices were used with position control for cursor manipulation but with rate control for navigation to avoid drawbacks of clutching operations for distance travel.
- While the tracking device's shape represents a certain orientation, the handles of the Bone on the other hand are close to have a spherical shape and therefore do not provide orientation cues.
- We used touch sensors for getting the information if the controllers of the Bone were grabbed or released. For the tracking devices a button had to be pressed to engage certain modes since the tracked handles could not be released during operation.

Raw input data of all applied sensor devices was filtered using the PRISM algorithm, presented by Frees and Kessler [FK05].

4. User Study

Our evaluation framework consists of four successive experiments, each emphasising on different interaction aspects:

- Experiment 1: selection and manipulation
- Experiment 2: flying / exploration
- Experiment 3: manoeuvring
- Experiment 4: combined interaction

Participants: We invited four students from our department to participate in our study. All of them had experience with 3D-graphics applications using common desktop devices. They did not have much experience with 3D tracking devices or game controllers. The small number of subjects is due to the long duration of the study.

Physical setup: For the he study we used a passive stereoscopic rear projection wall setup with 60 Hz frame rate. The projected image measured 2 m in height 3 m in width (Figure 6). The users were situated 2.5 m in front of the screen centre, either sitting (Bone) or standing (both devices). The physical size of the interactive objects was about one meter in diameter.

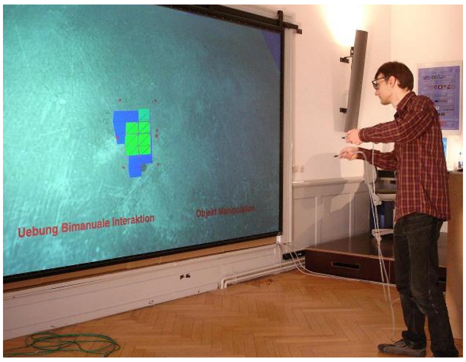
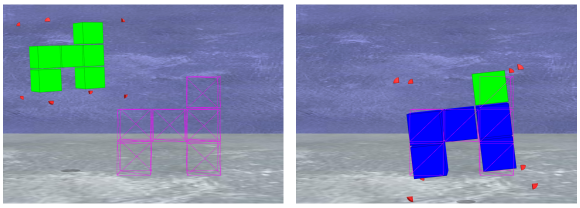


Figure 6: Physical setup (here shown in mono).

Design and Procedure: Each experiment started with a training session lasting no longer than 10 minutes, followed by 8 to 56 recorded trials depending on the experiment type. One of the sessions consisted of all the four experiments, lasting up to two hours. To minimize the effects of fatigue our participants performed only one session per day. For each device sessions took place during four consecutive days. On the first day with a new device training periods were extended to up to 30 minutes increasing total session time to up to three hours. The order of the devices was counterbalanced between subjects. After completing the eight sessions a written questionnaire had to be filled out, containing subjective usability and preference ratings.

4.1. Experiment 1: selection and manipulation

For each trial a new docking object was presented to the user, varying in shape, location and orientation. All shapes were simple combinations of cube primitives. For selection the described 3D-cursor had to be moved inside the object's geometry using three translational DOF of the controller on the dominant hand. A button click confirmed the selection. Now the docking task itself started by highlighting the corresponding docking position as a wire frame duplicate of the selected object (Figure 7a). The docking task afforded the usage of all six DOF. The requested accuracy was less than 5° for the rotations and 6 cm for the translations. A sufficiently close object position/orientation to the the target was visualized by change of object colour (Figure 7b). The object was deselected at the end of the task by another click on the cursor controller. Eight variations of docking setups were implemented (Table 1) containing various types of difficulties regarding number and type of degrees of freedom as well as magnitude of translational or angular distances per DOF. Each configuration is repeated seven times resulting in 56 recorded docking trials per session.



(a) Highlighting docking position. **(b)** Partially docked object.

Figure 7: Task setup for object selection and manipulation.

As proposed by [Zha95], the isotonic tracking devices were used with position control; the elastic controllers of the Bone were used with rate control instead.

Results: The Polhemus tracking device significantly outperforms the Bone (Figure 8) in both subtasks of this experiment, for selection ($p = 0.025$) as well as for the docking subtask ($p < 0.01$).

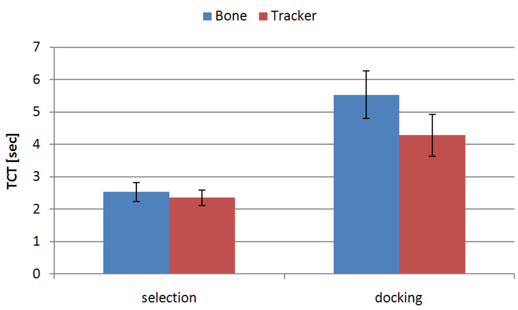


Figure 8: Mean TCT, STD for selection and docking subtask.

The slight performance benefit of position control for object manipulation confirms previous research [ZM98] and might decrease with more rate control practice. But in correspondence with [MBS97] we suggest, that proprioception is an important factor for manipulation performance. While rate control is solely based on a visual control loop to monitor the current position of manipulated objects, position control can additionally make use of proprioceptive feedback of executed movements which may allow for faster movements.

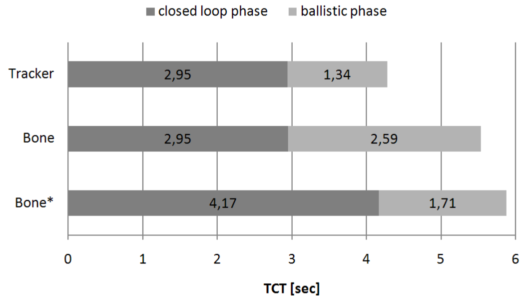
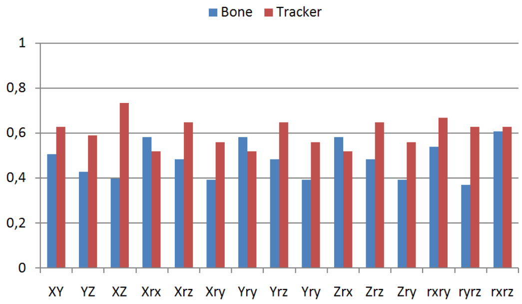


Figure 9: Breakdown of docking subtask TCTs into coarse grained ballistic phase movement and precise closed loop interaction for Tracker, Bone and high control display gain Bone mapping (Bone*).

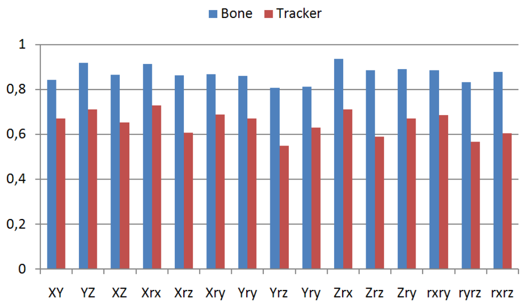
A closer analysis of the recorded data revealed that the overall performance advantage of the tracker is indeed based on faster coarse grained ballistic movements (Figure 9). The mean velocity of the position-controlled interface is three times as high as the mean velocity of the rate-controlled device. On average, it took *1.34 sec* with the tracker and *2.6 sec* with the bone to arrive in proximity of the docking position. During the final closed loop phase, both devices required the same amount of time (*2.95 sec*) for successful alignment, but differed with respect to interaction characteristics. While the Bone actions during the closed loop phase were slow and precise the tracker was faster but partially overshooting.

We arranged an additional experiment using a three times higher control display gain for the Bone (Bone* in Figure 9) to match the maximum observed velocity for the Tracker. In this case the Bone performed faster during the ballistic phase (*1.7sec*) but did not reach the Tracker's performance. Higher maximum velocity resulted in decreased precision during closed loop interaction: the Bone required much more time for the closed loop phase (*4.2sec*). This resulted in overall longer TCTs than with the originally chosen parameters. The characteristic of the Bone's closed loop phase now showed the same overshoot effects as found for the Tracker emphasising that a well balanced transfer function supporting the controversial requirements of both interaction phases is needed.

Masliah [MM00] introduced the M-metric, which allows a profound analysis of the use of multiple DOF. It computes the simultaneity of the used DOFs based on the fraction and magnitude of simultaneous operation as well as the effectiveness of the resulting motion through comparison of the optimal and the actually performed path. Using this tool, further differences between the operation of both devices became visible.



(a) Simultaneity: chronological and magnitudial congruence of multiple DOFs; 1 = fully simultaneous use of several DOF.



(b) Efficiency: ratio of optimal and recorded path of DOF combinations; 1 = recorded and path optimal identical.

Figure 10: M-Metric simultaneity (a) and efficiency (b) scores for two-way combinations of either translations (e. g. XY), rotations (e. g. rxry) or combinations of both motions (e. g. Xrx).

The mean simultaneity value (Figure 10a) for all DOF combinations for the Bone is *0.48*, for the tracker it is *0.6*, indicating that more degrees of freedom are used simultaneously with the tracker. However the path efficiency (Figure 10b) for the Bone is *0.8*, for the tracker it is *0.65*, which matches with the observed slower but more precise Bone operations. It also shows that the observed overshoot with the tracker lead to a longer and less efficient path even though the overall task performance was better.

Figure 11 shows that the Tracker performance was quite similar throughout different task difficulties described in Table 1. The TCT of the simplest case *A* requiring two DOF was identical to the more demanding cases *B*, *D*, *E*, *F* and *H* where up to five DOF had to be controlled. The TCTs of the Tracker increased if the magnitude of translations and rotations exceeded the physically available manipulation range of the human hand-arm system. For these cases, the position tracker must be clutched, which is not only time consuming but also annoying. The elastic rate controllers never require clutching and therefore catch up in performance for these tasks (cases *C* and *G*). The TCT of the Bone depends on the employed DOFs. For the pure translational settings of case *A* and *E* requiring a different number of DOF, TCTs were equal, but if translations and rotations are concurrently required, TCTs increased significantly. There was a particular dependence on the amount of rotations required, which indicates that the isotonic Tracker rotations are more efficient.

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
translation	2	2	2	2	3	3	3*	3
rotation	0	1	1*	2	0	1	0	2

Table 1: DOF combinations of used docking setups. DOF signed with * consist of three times higher magnitude than default ones.

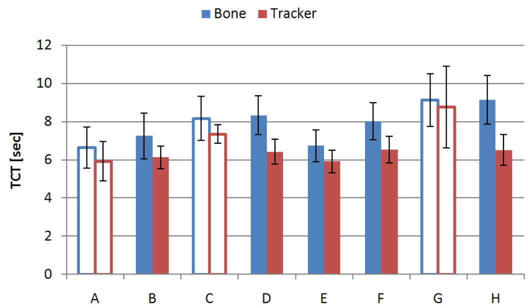


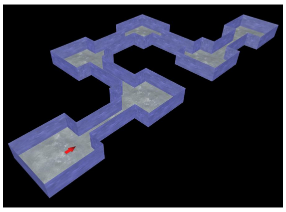
Figure 11: Mean TCT, STD for various task complexities. No significant differences between devices in highlighted cases.

The questionnaire revealed that subjects strongly preferred position-controlled tracking input for object manipulation over rate-controlled input even though the task performance was only slightly different indicating that it might require more cognitive efforts to achieve similar performance.

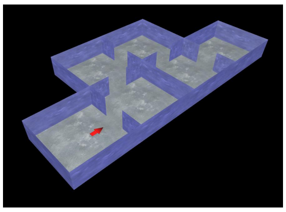
4.2. Experiment 2: flying/ exploration

A maze-like travel path containing a number of rooms and connecting corridors was presented to the participants. The users were requested to navigate from room to room without hitting the walls. The further movement direction was represented by an arrow on the floor, if the subject got in close proximity to the next room in sequence. We tested four different paths, each emphasising on particular travel aspects. During a session all of these paths had to be passed in both directions. Path *A* alternates $\pm 90°$ corners and straight lines (Figure 12a) while path *B* featured the same curvature setup without the straight line segments (Figure 12b). Path *C* was composed of vertical ramps, directing up and down (Figure 12c). Path *D* combined all the aforementioned aspects (Figure 12d).

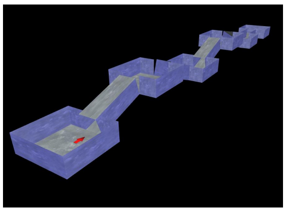
Results: For this experiment we employed rate control for the Tracker and the Bone. We expected better performance for the Bone than for the Tracker since, according to Zhai [Zha95], isotonic devices are inferior to elastic devices for rate-controlled object manipulation tasks. However our task was navigation and our results do not show any significant difference among the devices ($p = 0.479$). This is probably due to reduced precision requirements for long distance travel in comparison to object manipulation tasks. Nevertheless, participants preferred the Bone for this task.



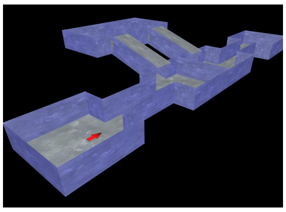
(a) Forward and head movement. Straight segments in between.



(b) Forward and head movement. Without straight segments.



(c) Forward and pitch movement.



(d) Forward, head and pitch.

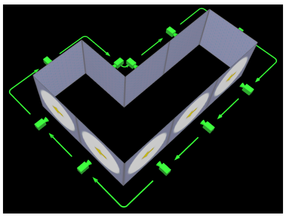
Figure 12: Employed path setups for exploration task.

A detailed analysis of the recorded data shows that the way subjects are steering during travel differs a lot. With the Bone, they suddenly adapted to the viewpoint-cursor constraint of the implemented interaction technique and tended to induce viewpoint rotation through translational input as described in chapter 3.4. About *37.4%* of the trial time subjects assign both controllers to different subtasks, while only *9.1%* of the total trial time both controllers were used to induce rotations symmetrically. The two Trackers on the other hand were used in a symmetrical manner such that for viewpoint rotations participants made use of the integral degrees of freedom of the devices. For *53%* of the total trial time both tracking devices were used symmetrically for steering. We note, that participants tended to hold both trackers close together and handle them like a single input device during distance travel.

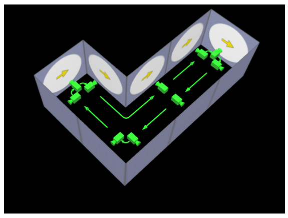
Whether employing both controllers symmetrically for travelling as with the Tracker or distributing heading and thrust as subtasks to both hands as with the Bone does not seem to affect task performance. Heading and thrust are obviously integral attributes of travelling. Thus the integral attributes of a task do not need to be controlled just by one hand. They may as well be operated by both hands simultaneously, which could be considered a generalization of Jacob et al.'s theory [JSPF94] that integral attributes of a task should be compatible with the simultaneously available DOF.

4.3. Experiment 3: manoeuvring

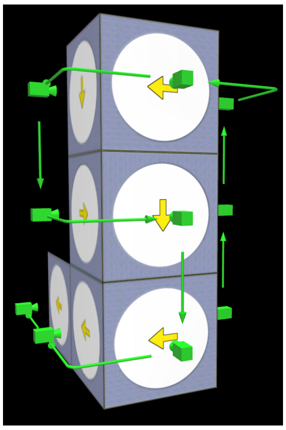
This experiment required exact viewpoint adjustment to a given surface patch. Subjects had to manoeuvre in close proximity and perpendicular orientation towards multiple given patches of an L-shaped reference object. The accepted orientation mismatch was less than 20° (angle between viewing vector and segment normal) while maximum distance between viewpoint and segment was set to 1m. Fulfilling these conditions initiated the appearance of the next segment in line. Four different manoeuvring setups were conducted, featuring different kinds of predefined paths to view all target patches of an object. Besides a horizontal setup (Figure 13a), requiring three DOF (depth, horizontal strafes, heading) to be controlled, a more complex setup (Figure 13c) with up to five DOF (depth, horizontal and vertical strafes, head and pitch) was used for the study. Each path setup was executed in outside-looking-in (Figure 13a, Figure 13c) and inside-looking-out manner (Figure 13b, Figure 13d). The outside-looking-in type focused on strafing operations (straight and curved) while inside-looking-out particularly required egocentric view browsing. To avoid an impact of cognitive way finding capabilities on the recorded data, an arrow pointed towards the next target patch. The sign appeared on the actual surface segment, once the view was correctly aligned.



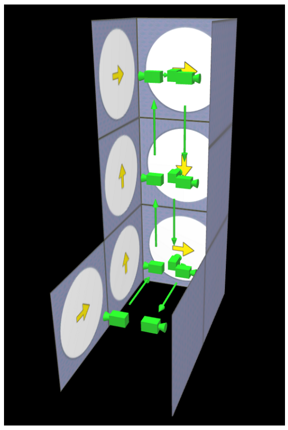
(a) Horizontal strafe and head. Outside-looking-in manner.



(b) Horizontal strafe and head. Inside-looking-out manner.



(c) Horizontal and vertical strafe, head and pitch. Outside-looking-in manner.



(d) Horizontal and vertical strafe, head and pitch. Inside-looking-out manner.

Figure 13: Viewpoint adjustment to given object patches. Path of viewpoint movement symbolized by green cameras and arrows.

Results: In contrast to the exploration task results, manoeuvring capabilities of the devices differ from each other (Figure 14). The reason for the Bone's significant advantage ($p < 0.01$, $p = 0.09$, $p < 0.01$ and $p < 0.01$ for paths *A - D*) seems to be the intensive use (*25%* of TCT) of the camera-cursor connection metaphor. Circling around corners is often required during manoeuvering tasks. With the Bone this is done through coordinated translational input of camera and cursor controller, which results in encompassing movement arround each other. With the two Trackers this kind of reference-based interaction is not utilized at all (*3%* of TCT). The fraction of uni-manual interaction with the tracking system increased instead to *70%* compared to only *5%* during the exploration experiment.

	A	*B*	*C*	*D*
trans	depth sideward	depth sideward	depth sideward upward	depth sideward upward
rot	head	head	head pitch	head pitch
view	out-in	in-out	out-in	in-out

Table 2: DOF combinations of used manoeuvring setups.

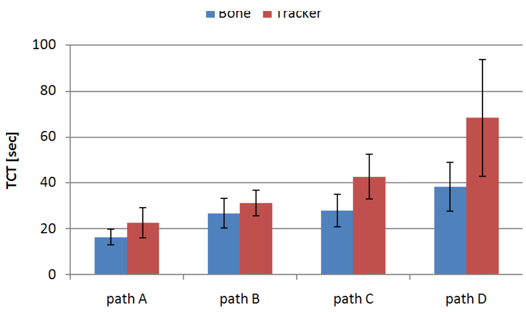


Figure 14: Mean TCT, STD for different manoeuvring paths.

Figure 14 illustrates the performance of both interaction systems in relation to task difficulty (Table 2). In contrast to the results of the manipulation experiment, where the tracker performance was not much influenced by the number of required DOF, here it is. For rate-controlled navigation tasks, the elastic controllers, operated with both hands simultaneously to adjust five DOF, show less dependency on the task difficulty than the tracking solution, which was operated mostly one-handed, using the integral DOF of one sensor.

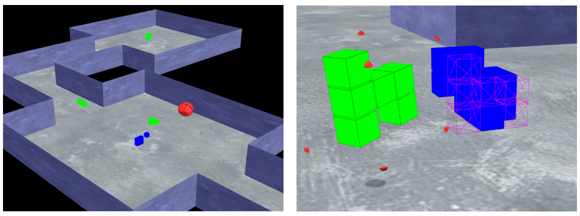
4.4. Experiment 4: combined interaction

After testing the usability of the proposed interface techniques in isolated subtasks, subjects were finally asked to combine the exercised skills in a more demanding task.

The virtual environment presented to the subjects was a simple architecture containing one large and two smaller rooms on a single floor. With the large room to the midst, they were connected through L-shaped corridors (Figure 15a). A 3D-puzzle marked the docking area in the centre of the large room. Initially positioned near that centre, subjects were instructed to travel to one of the neighbouring rooms by following the arrows as known form experiment two. There they found one of these simple-shaped docking objects known from experiment one, which they had to pick up. Carrying the object, participants were then requested to turn back to the docking area where the object had to be inserted into the 3D Puzzle. The target position itself was partially hidden by other geometric objects (Figure 15b), which forced manoeuvring operations. Further docking parameters remained the same as described for experiment one.

For object transport over distance, the manipulated object could be stored inside a virtual backpack. To do so, the manipulated object could be directly moved across the viewing plane level and dropped there. Alternatively a double click on the cursor controller served as a shortcut for this function to unblock the 3D-cursor for bimanual navigation. Unpacking worked in both ways similarly.

The combined interaction trial contained four successive phases: it started with a distance travel task, followed by a selection task, then object transport over distance and finally an extended docking task that involved manoeuvring. To ensure that subjects would not adopt certain interaction habits, but employ different strategies, six different configurations were performed three times per session. While the task difficulties and overall movement distances remained similar, varying docking objects and associated target positions were applied.



(a) Distance travel overview. **(b)** Extended docking task.

Figure 15: Employed setup for combined interaction task consisting of travel, manoeuvring, selection, manipulation and object transport.

Results: This task required an equal amount of navigation and manipulation. Based on the results of the previous experiments, we expected similar task performance for Tracker and Bone since the advantages and disadvantages of both devices should even out. In fact the TCTs turned out to be quite similar: *28.6 sec* were observed for the Bone and *27.3 sec* for the Tracker. The statistical analysis revealed no significant difference ($p = 0.451$) indicating that both devices performed on a similar level.

5. Conclusion and Future Work

The comparison of tracking- and controller-based interaction shows that the intuitiveness and efficiency of isotonic position control for 3D object manipulation can hardly be achieved with elastic rate control devices. However, elastic controller-based input resulted in better performance for rate-controlled navigation. The simultaneous operation of multiple DOF, which is beneficial for our tasks, was easier to achieve with the tracked handle than with the elastic controller-based device. We realized that our asymmetric two-handed navigation technique, which facilitated complex manoeuvres, was much less used with the tracking devices than with the Bone. This indicates that it is cognitively more difficult to keep track of the position and orientation of both hands than it is to operate two separate elastic sensors with the fingertips of both hands. Besides the elastic force feedback, the major reason might be the assembly of both sensors in a single housing, which serves as a common reference for both hands' input.

We found that even for compound interaction tasks, controller-based handheld devices bear the potential to compete with tracking-based input systems commonly employed for Virtual Reality applications. In addition, our study shows that control sensors manipulated with the fingertips may work better than tracking devices for certain tasks. In our case manoeuvring could be performed better with the controller-based device. It is very likely that other sensor configurations would increase performance in another field of 3D interaction since input sensors strongly vary in shape, functionality and size. However, the combination of tracking-based and controller-based input in a single device could provide the best of both worlds.

The trend of entertainment media toward more digitally created 3D content is apparent, which requires intuitive and ergonomic input devices that may be used while sitting on the sofa or standing in front of a large screen TV. Another example are 3D digitized sports events, which allow users to fly around baseball fields or to directly follow the ball of a broadcasted soccer game. Controller-based multi-DOF input devices such as the Bone have the potential to become widely used for these application domains if appropriate transfer functions and interaction metaphors are developed.

References

[BJH99] BOWMAN D., JOHNSON D., HODGES L.: Testbed Evaluation of Virtual Environment Interaction Techniques. In *VRST '99* (1999), pp. 26–33.

[BK99] BALAKRISHNAN R., KURTENBACH G.: Exploring Bimanual Camera Control and Object Manipulation in 3D Graphics Interfaces. In *Proc. of CHI '99* (1999), pp. 56–62.

[BKVP05] BOWMAN D., KRUJIFF E., VIOLA J. L., POUPYREV I.: *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2005.

[BM86] BUXTON W., MYERS B. A.: A study in two-handed input. In *Proc. of CHI '86* (1986), pp. 321–326.

[Bow99] BOWMAN D.: Interaction Techniques for Common Task in Immersive Virtual Environments. In *PhD-thesis* (1999).

[CMN80] CARD S., MORAN T., NEWELL A.: The keystroke-level model for user performance time with interactive systems. In *Communications of the ACM Vol. 23 7* (1980), pp. 396–410.

[FK05] FREES S., KESSLER G. D.: Precise and Rapid Interaction through Scaled Manipulation. In *Proc. of VR '05* (2005), pp. 99–106.

[Gar74] GARNER W. R.: The Processing of Information and Structure. In *Lawrence Erlbaum Associates* (1974).

[Gui87] GUIARD Y.: Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. In *Journal of Motor Behavior 19* (1987), pp. 486–517.

[HPP*97] HINCKLEY K., PAUSCH R., PROFFITT D., PATTON J., KASSEL N.: Cooperative Bimanual Action. In *Proc. of CHI '97* (1997), pp. 27–34.

[HSK*05] HUCKAUF A., SPEED A., KUNERT A., HOCHSTRATE J., FRÖHLICH B.: Evaluation of 12DOF Input Devices for Navigation and Manipulation in Virtual Environments. In *Proc. INTERACT '05* (2005), pp. 601–614.

[JSPF94] JACOB R., SILBERT R., PRESTON M., FARLANE D. M.: Integrality and Separability of Input Devices. In *Proc. of CHI '94* (1994), pp. 3–26.

[LZB99] LEGANCHUCK A., ZHAI S., BUXTON W.: Manual and Cognitive Benefits of two handed Input: an Experimental Study. In *Proc. of CHI '99* (1999), pp. 326–359.

[MBS97] MINE M., BROOKS F., SEQUIN C.: Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction. In *Proc. of SIGGRAPH '97* (1997), pp. 19–26.

[MM00] MASLIAH M., MILGRAM P.: Meassuring the Allocation of Control in a 6 Degree-of-Freedom Docking Experiment. In *Proc. of CHI '00* (2000), pp. 25–32.

[SF03] SIMON A., FRÖHLICH B.: The YOYO: A Hendheld Device Combining Elastic and Isotonic Input. In *Proc. of INTERACT '03* (2003), pp. 303–310.

[Sim74] SIMON H.: How Big is a Chunk. In *Science 183* (1974), pp. 482–488.

[ZBM94] ZHAI S., BUXTON W., MILGRAM P.: The "Silk Cursor": investigating transparency for 3D target acquisition. In *Proc. of SIGCHI '94* (1994), pp. 459–464.

[ZFS97] ZELEZNIK R., FORSBERG A., STRAUSS P.: Two Pointer Input for 3D Interaction. In *Proc. of SI3D '97* (1997), pp. 115–ff.

[Zha95] ZHAI S.: Human Performance in Six Degree of Freedom Input Control. In *PhD-thesis* (1995).

[ZM98] ZHAI S., MILGRAM P.: Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In *Proc. of SIGCHI '98* (1998), pp. 320–327.