On Motor Performance in Virtual 3D Object Manipulation

Alexander Kulik*   André Kunert*   Bernd Froehlich*
Virtual Reality and Visualization Research, Bauhaus-Universität Weimar

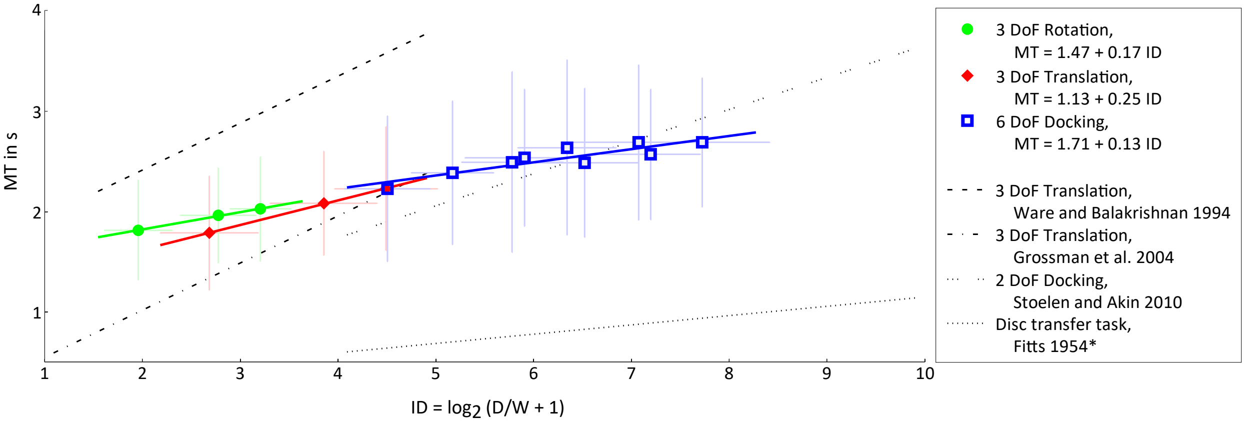


Fig. 1: Linear regression with Fitts's model of motor performance in three different aimed movement tasks: 3D Rotation, 3D Translation, and 3D Docking with six degrees of freedom. Indices of difficulty were derived from effective target widths and distances, hence, varying for each user and subtask (error bars show standard deviations on both axes). For comparison, we also plotted the linear regressions from prior studies [12, 20, 57, 69] in the range of the respectively tested IDs.

*Fitts's original data from the disc transfer task recomputed by MacKenzie [39] using the Shannon formulation

Abstract—Fitts's law facilitates approximate comparisons of target acquisition performance across a variety of settings. Conceptually, also the index of difficulty of 3D object manipulation with six degrees of freedom can be computed, which allows the comparison of results from different studies. Prior experiments, however, often revealed much worse performance than one would reasonably expect on this basis. We argue that this discrepancy stems from confounding variables and show how Fitts's law and related research methods can be applied to isolate and identify relevant factors of motor performance in 3D manipulation tasks. The results of a formal user study (n=21) demonstrate competitive performance in compliance with Fitts's model and provide empirical evidence that simultaneous 3D rotation and translation can be beneficial.

Index Terms—Fitts's law, throughput, motor performance, aimed movements, 3D user interfaces

◆

1 INTRODUCTION

3D user interfaces and interaction techniques are often evaluated in terms of user performance in object docking tasks with six degrees of freedom (DoF). The involved motor actions are aimed movements that require users to find an optimal tradeoff between rapidity and accuracy. Compliance with Fitts's law can thus be expected which would facilitate comparisons of motor performance across different studies and even across different types of aimed movement tasks. To our knowledge, however, Fitts's model had not yet been used to establish reference performance measures for virtual 3D object manipulation.

We reviewed literature on various aimed movement studies to identify benchmark performance measures for canonical 2D and 3D selection and object manipulation tasks. The latter are typically not modeled by Fitts's law, but estimates of the tested indices of difficulty can be derived from the given task parameters (Sec. Sect. 2.4). User performance often appeared to be comparably low in this case. Accuracy limits of human motor control and fatigue are possible reasons since physical support and tactile feedback are lacking in mid air interactions, but other possible explanations include cognitive load and particular interaction overheads of the respective task and apparatus.

To minimize the effects of such confouding variables, we devised a novel experimental approach for research on 3D object manipulation and applied it in a formal user study ($n = 21$). The results may serve as reference performance measures for virtual 3D object manipulation with six DoF (Sec. Sect. 3.6). Our main goal was to enable comparisons of motor performance in virtual 3D object manipulation with other aimed movements, e.g. 2D pointing. We believe that this approach can lead to a better understanding of the benefits and drawbacks of direct 3D manipulation in comparison to more constrained interaction techniques.

We were motivated by the fact that Fitts's law had already been validated for various types of movement (Sec. Sect. 2). Its application to studies on virtual 3D object manipulation may thus facilitate the desired comparability with prior research, e.g., to validate the implementation quality of established interaction techniques (often used for comparisons) and to substantiate the results of user studies with typically only a few participants. The main contributions of our work are:

- a novel approach to study motor performance in 3D object manipulation with reciprocal target acquisition tasks,
- empirical support that motor performance in 3D object manipulation tasks can be modeled by Fitts's law, and
- empirical evidence for benefits of combining translation and rotation input.

We also provide a literature review on benchmark performance in different aimed movement studies and the results of our study offer conservative reference measures for virtual 3D object manipulation.

---

*e-mail: <firstname>.<lastname>@uni-weimar.de

Our experimental paradigm allowed participants to find their individually preferred balance between rapidity and accuracy. This reduced potential effects of individual movement precision but it also led to a comparably long closed-loop phase during which users evaluated and confirmed the achieved accuracy. As a side effect, the obtained measures of accuracy may also serve as indicators of practically achievable placement accuracy.

2 RELATED WORK

Graphical workstations are commonly operated with 2D input devices, even for the specification of 3D transformations, e.g. in computer aided design applications (CAD). It was argued that 3D input techniques are generally not competitive to more constrained methods like their 2D counterpart [4, 43, 58]) Other researchers found the contrary [55] or demonstrated more specifically, that integral 3D motion input is faster but less accurate than its decomposition into 2D input operations [63]) The potential benefits of integral 3D translations and integral 3D rotations are, quite obviously, related to the shorter diagonal motion path across all three dimensions, but they may be outweighed by other factors. More consecutive research is necessary to gain a better understanding.

There is even less agreement over potential beneficial of simultaneous rotation and translation [16, 41, 42, 66]. In fact, the results of most studies on object docking with six degrees of freedom in virtual environments reveal surprisingly long task completion times [9, 64, 67, 74], while similar actions in real-world settings seem to be competitive to user performance known from pointing studies [65, 66]. The source of such differences is unknown.

Chapoulie et al., for example, compared user performance in aimed movement tasks with different degrees of freedom (DoF) between real world settings and virtual reality [9]. They included rotation and translation tasks (1D and 3D respectively) as well as their combination. Special-purpose hardware devices were built to enable fair comparisons between real and virtual settings. The results confirmed benefits of tangible interaction with physical objects as known from prior studies [19, 47]. Unfortunately, the implementation of the different tasks with varying degrees of freedom, did not follow the same paradigm to specify the requested target region. Thus, their difficulties cannot be directly compared, but it is surprising that average task completion times of 20 to 33 seconds were recorded for placing objects within reach. It seems plausible that these results are much stronger related to differences of the provided visual feedback than to motor performance.

We argue that Fitts's law offers a viable model for all these types of aimed movements and its consistent application would facilitate more reliable comparisons between different experimental tasks and independent studies. Fitts developed his model with 3D assembly processes in mind and it has been successfully applied to real and virtual object manipulation (e.g. [24, 31]). His experimental paradigm, using reciprocal movements between two target locations, nicely isolates the motor effort from cognitive processes of motion planning. However, it has not been considered so far for studies on virtual 3D object manipulation with six degrees of freedom.

2.1 Performance Measures of Rapid Aimed Movement

The operation of direct manipulation interfaces can be broken down into sequences of target acquisitions in space. Such aimed movements, in turn, can be modelled with Fitts's law[1] [12]: a linear regression model with the intercept a and the slope coefficient b that relates the required movement time (MT) to the task's index of difficulty (ID). The latter is defined as the binary logarithm of the ratio between the amplitude and the accuracy of the movement (equation 1).

The logarithmic relation reflects that movements towards a known target involve a ballistic phase for coarse approximation. Further away targets are approached with a higher velocity, and thus the movement time does not increase linearly with the target distance. The smaller the target, the earlier this open-loop process must give way for closed-loop control with higher motion accuracy[2].

The formulation of Fitts's index of difficulty (ID) is disputed in the community [11]. Its original form reflects that the acquisition tolerance is only half the target width: $ID = log_2(2D/W)$ [12]. However, the Shannon formulation, suggested by MacKenzie [40] is more popular. IDs computed with the latter are about 1 bit lower than those computed with the original formula ($log_2(2) = 1$). We used the Shannon formulation for all computations of ID in this paper to facilitate comparisons across different studies:

$$ID = log_2(D/W + 1) \tag{1}$$

The robustness of this model has been demonstrated for translational [8, 20, 28, 37, 45, 54], rotational [9, 44, 57] and even scaling movements [61]. Stoelen suggested that also combinations of rotation and translation can be modeled using the sum of the indices of difficulty of both submovements (equation 2 [57]). This approach reflects that rotation and translation are independent spatial transformations.

$$MT = a + b(ID_{rot} + ID_{trans}) \tag{2}$$

Fitts's model facilitates the comparison of various interfaces, if additional factors of user performance are otherwise controlled. The skills of test users or distractions, for example, are not expressed in Fitts's law parameters. The performance in aimed movements also depends on strategy. More than 30 % shorter task completion times have been observed if rapidity was favored over accuracy [37]. The side effect of such rapidity are larger deviations from the actual motion endpoint. MacKenzie suggested to compute the effective target width and distance from this actual distribution of hits which results in comparable throughput measures that are independent of user strategies [37, 39].

Two different approaches have been suggested for computing an overall performance measure. The index of performance or throughput (TP) is generally defined as the ratio of ID and MT [12]. It is thus affected by both, the intercept a and the slope coefficient b of the regression model. Zhai clarified that throughput comparisons on this basis are biased. The intercept a has a stronger effect for small IDs than for large ones [73]. More reliable comparisons across various IDs may thus be possible if only the reciprocal of the slope coefficient b is considered. This approach, however, obscures constant interaction overheads, e.g., systematic activation latency of the input sensor.

If the necessary data is provided, throughput measures enable the coarse comparison of experimental results with those of related studies. The interpretation must take several additional factors into account, e.g., the mentioned interaction between the intercept a and different IDs. Further sources of error include varying impact of target width and distance. In a study of Graham and MacKenzie, for example, target width had a slightly larger impact than distance [19]. Formally, such observations indicate a violation of Fitts's law. They may be an effect of the formula used for the computation of ID (see above) or, more likely, they stem from accuracy limitations of the human perception-action loop in combination with the respective experimental apparatus.

In order to test for the predicted scaling invariance, Guiard suggested to vary form (ID) and scale of experimental tasks in Fitts's law research instead of target distance and width [21]. However, he also noted that only few combinations of target distance and width, that can be comfortably reached by humans, yield the same IDs and concluded that varying only target distance while fixing target width at a comfortable level can be considered an alternative for practical applications.

Our motor skills also vary with the direction of movement. Upward movements, for example, seem to be more demanding than those downwards or to the side [8, 45], while even better performance can be achieved in forward movements [8]. Interestingly, such movements in depth seem to be more difficult with 3D computer interfaces [3, 20, 69]. This may be an effect of missing tactile feedback and/or impaired depth perception on a 3D display.

Several adaptations of Fitts's law were suggested to account for the observed effects [8, 19, 20, 45]. These extended models yield a better fit, but complicate comparisons between different settings. Fitts's law is an approximate model that ignores many factors of real-world interaction. The results of independent studies thus cannot be compared formally,

---

[1] see [40] and [73] for extensive introductions to Fitts's law in HCI

[2] see [35] and [47] for a detailed analysis of movement phases

but, they indicate a reasonable range of performance in fundamental target acquisition tasks. Results from various studies can be critically reviewed with respect to such ballpark figures.

Rough comparisons seem to be possible, even if the actual range of tested IDs is not reported. The logarithmic term makes extreme IDs unlikely if reasonable target sizes and distances are applied. For example, reaching a 2 centimeter wide target area at 6 centimeter distance, corresponds to an ID of 2, while the ID of aiming over 1 meter distance at a target of 2 millimeter width is slightly below 10. Fitts's law does not hold for extremely small targets [10] and extremely large distances. The latter can be more accurately modeled as a sequence of target acquisition tasks [7] or a combination with scaling [22, 23]. Researchers generally avoid such extreme cases, if they are not deliberately interested in their effects, hence IDs beyond 10 are not to be expected. Commonly used IDs are in the range of 3-9 [23].

Observations and insights from over half a century of research on the topic, demonstrate the broad applicability of Fitts's logarithmic model to various types of aimed movements. Stoelen and Akin [57] even suggested to model combinations of movement types like rotation and translation as the sum of their indices of difficulty. However, the rich available data had not been considered so far for coarse comparison of the achievable motor performance in these different settings.

2.2 Performance in Pointing and Object Translation Tasks

Fitts's original studies examined 1D pointing performance with styluses of different weight as well as 3D object transfer (pins and washers). He computed throughputs for the different conditions in the range of 10 to 11 with outliers for very small IDs. Effects of object weight could be observed. MacKenzie later recomputed these results using the Shannon formulation and the effective index of difficulty [39], which yielded slightly lower throughputs (8 to 9) for the stylus conditions, where participants could miss the accurate target location.

Fitts's original experiment involved repetitive acquisitions of the same two targets. A later study with random target locations revealed an increase of reaction time, which was necessary for motion planning [13]. The pure movement times in this study were 30-50 % shorter for comparable IDs, but if we take into account the almost constant reaction time of about 300 ms, the throughput in these discrete target acquisition tasks dropped to approximately 8. 2D pointing studies with computer interfaces like mice, touchscreens, or digitizing pens often resulted in lower throughputs of 4 to 6 [14, 37, 38, 54, 71, 72], hence it took about twice as long to perform comparable tasks. Graham and MacKenzie directly compared virtual and real pointing. They recorded throughput rates of about 9 for the real condition and 8.5 for the virtual condition [19]. Apparently, pointing performance can be comparable in both cases, but it is typically lower with computer interfaces.

For 3D pointing in real-world settings, Murata and Iwase reported an average throughput of about 4.7 [45]. From the data provided by Cha and Myung we computed an average of about 5. Studies on 3D pointing in virtual environments revealed throughput rates in the range of 2 to 2.5 [1, 20, 36, 59] or lower [69]. A notable exception is a recent study by Batmaz et al., where mean throughput levels above 4 could be achieved in lateral target acquisition tasks using VR and AR headsets [3]. The performance of 3D pointing at virtual targets seems to suffer primarily from latency [69], a lack of tactile feedback [35, 47, 59], and impaired depth perception [3, 36], which negatively affect accuracy.

The mentioned results from studies on pointing and pure translation tasks, indicate, that performance with computer interfaces is typically lower than it is in real world settings, but at least with well designed 2D user interfaces, the differences can become neglectable. 3D target acquisition performance, on the other hand, appears to be significantly more difficult than its 2D counterpart. A difference that can outweigh the benefits of simultanous error reduction. Movement accuracy seems to be the limiting factor here, rather than rapidity. It remains an open research question, if better performance can be achieved in settings with reduced accuracy requirements.

2.3 Performance in Rotation Tasks

3D object rotation can be very difficult. In principle, the optimal transition between any two orientations is a single turn that never exceeds 180 °. However, the axis of rotation is not intuitively clear if it does not correspond to any of the principal axes of the object or the environment. Parsons showed that in this case people have a hard time to tell whether two objects in different orientations are of the same shape or to describe the shortest motion path between both postures in terms of rotation angle and axis [49]. If the target orientation cannot be identified before the motion onset, the applied rotation is not an aimed movement, but rather a search process. Note that this is a rare case in the handling of everyday objects.

Active object manipulation alleviates the difficulty of mental rotations. By trying we can rapidly align two objects and compare their morphological similarity. Zhai and Milgram showed that any 3D rotation task can be successfully solved in interactive settings, although with lower efficiency in case of oblique rotation axes [74]. User performance in rotation tasks seems to depend strongly on the geometric complexity of the rotation object and the orientation of the rotation axis. When comparing rotation performance across different tasks and settings we must expect a large variance. Nevertheless, best-case results from the literature may serve as benchmarks.

Fitts's law has been shown to apply in cases where the angle and axis of the rotation is obvious and constrained. Meyer et al. [44], for example, used a mechanical handle with one DoF that required wrist rotation similar to turning a key. They also suggested a square root function for ID instead of Fitts's logarithmic model, to express a slightly different conceptual model for rapid movements[3], but noted that it makes only a marginal numerical difference for their test conditions. Average movement times for rotations with angular target distances between 10 and 40 ° and target widths in the range of 1.6-6.3 ° were no longer than 0.3-0.6 s. Their participants tended to be less accurate than required for small targets and more accurate for large targets, hence, they computed subjective target widths in the range of 2.0-5.7 °. We recomputed the tested indices of difficulty using the Shannon formulation and derived throughput rates in the range of 5.5-6.8. Similar results were obtained in a follow-up study without visual feedback.

In another study by Stoelen and Akin [57], only the visual stimulus on a screen was constrained to one rotational DoF, while the input device, a wired electromagnetic 3D tracker, could also be translated during its operation on a tabletop. They obtained average movement times of 0.7 s to 1.75 s for rotation tasks with indices of diffculty between 1.5 and 5.0, which corresponds to throughput rates of 2.1 to 2.9.

Kopper et al. [32] argued that distant pointing at large screens, can be more accurately modeled as rotations of the input device rather than with linear target distances and widths on the screen. They also observed that accuracy is often more demanding than rapidity in this case and thus incorporated two exponential weighting factors to their model of ID. This adaptation contradicts the assumption of scaling invariance in Fitts's original model, but it offered a better model fit to their data. However, it also impedes rough performance comparison with the results of other studies.

Ware and Rose compared 3D rotations of real and virtual objects [70]. Their results demonstrated that rotating real objects about 125 ° on average with a tolerance of 4-5 ° can be achieved in less than two seconds [70]. This corresponds to a throughput of roughly 2.6. In a blindfolded condition it took the participants about 17 % longer to reach the target rotation with an average error of about 9.21 °. If the same task was performed with the same physical handles as input, but also with visual feedback from computer graphics, participants could achieve a similar accuracy as in the real-world condition, but it took them 25 % longer. Ware and Rose suggested that the reason for the slightly lower performance was the system latency of about 75 milliseconds. Other studies on 3D rotation of virtual objects generally observed worse performance (e.g. [27, 34, 51, 68, 75]).

---

[3] Their *stochastic optimized-submovement model* assumes more variability for the endpoint and duration of each submovement than the earlier suggested *deterministic iterative-corrections model*.

Rotational aimed movements were studied much less than their translational counterpart, but compliance with Fitts's law has been demonstrated multiple times–at least, if challenging mental rotations can be avoided. The available data from prior work, indicates lower motor performance in rotational than translational movements. Further experiments under comparable conditions are needed to better understand how both movement types compare with each other. A related question is, if and to what extend, oblique object orientations affect user performance in reciprocal 3D rotation tasks. The cognitive effort can be expected to vanish after a few repetitions in this case.

2.4 Performance in Object Manipulation Tasks

So far we have considered user performance in purely translational and purely rotational tasks. Moving objects in space involves simultaneous 3D rotation and translation with six DoF. Docking tasks that require the 3D alignment of two equally shaped objects are typically applied for corresponding performance studies (e.g. [33, 43, 64, 74]). It is an open debate, whether the simultaneous control of rotation and translation is efficient, but it is the commonly observed behavior during the manipulation of real and virtual objects [33, 57, 66, 67].

Masliah and Milgram showed that this simultaneity is not necessarily more efficient than separate operations [42]. The index of difficulty in 6-DoF manipulation tasks might thus be adequately defined as the sum of two aimed movements that describe both submovements separately. Stoelen and Akin found a good fit for this simple model, although the participants of their study operated rotation and translation simultaneously (all but one of 13) [57]. We follow their example and use this combined index of difficulty to compare the performance of aimed movements with more DoF. It is a conservative estimate which neglects potential benefits from simultaneous rotation and translation.

The results of Stoelen and Akin on user performance in a reciprocal target acquisition task that combined rotation and translation (each only 1D) may serve as a reference for simple object docking in computer applications. For combined IDs in the range of 5 to 12, they obtained task completion times between 2 and 3.5 seconds which corresponds to a throughput between 2.5 and 3.5. Wang et al. earlier studied a similar task with a wooden cube on a physical table [65, 66]. Based on the reported data, an effective index of difficulty of about 6 can be computed. The obtained task completion times of 776 milliseconds on average thus correspond to an effective throughput of roughly 7.6. This performance is almost two times better than that obtained by Stoelen and Akin for virtual object manipulation but it compares quite well to the pointing performance obtained in Fitts's original experiments.

Experiments on docking virtual objects with more degrees of freedom often revealed worse performance. Ware reported 14.05 seconds for a 6-DoF docking task with isotonic position control [67]. We can estimate a combined ID of about 8 (using the average error for target width), hence the throughput would be in the range of 0.6. Zhai and Milgram showed that users can strongly improve 3D docking performance with training and reduce the task completion times from almost 20 seconds to less than 7 seconds. Their descriptions of tasks and the results does not support the computation of an index of difficulty, but under consideration of their experimental apparatus we do not assume higher IDs than those studied by Ware [74]. In more recent studies the recorded task completion times were in the range of 15-20 seconds [64] or higher [9, 43], but again, the reported data does not allow the computation of task difficulties. We note, that the target locations in both cases were within arms reach, hence, very high IDs are unlikely.

Prior experiments on 3D object manipulation often resulted in surprisingly long task completion times, that do not compare well to those obtained in aimed movement studies with only translational or rotational DoF. Stoelen and Akin's suggestion to model the combination of both as the sum of their IDs offers a basis for more meaningful comparisons, but even from this perspective, the results are not convincing. If the reasons for overall low motor performance are unclear, it seems impossible to identify the decisive factors in comparative evaluations of novel user interfaces and interaction techniques. This observation calls for a novel experimental paradigm to isolate motor performance from confounding variables in studies on 3D object manipulation. Stoelen and Akin's reciprocal target acquisition task seems to be a good candidate, but it was constrained to only one degree of freedom per movement type. Further research is necessary to establish generic experimental methods for tasks with varying DoF and to derive robust reference measures for rough comparisons between the results of independent studies.

2.5 Integrality and Separability of Rotation and Translation

Research on interfaces for rotation and translation was often informed by the concept of perceptually integral or separable task attributes. Jacob et al. suggested that the control structure of input devices should reflect the corresponding perceptual structure [30]. Wang et al. argued that "..., object transportation and orientation could be integrable because the spatial attributes are generally considered integral" [66]. Martinet et al. shared this intuition [41]. Research on cognitive information processing, instead, indicates that position, orientation and distance are perceived and processed separately [17, p.134] [29], while variations in vertical and horizontal position are processed integrally [18]. However, the concept of perceptual integrality and separability is not a dichotomy that would support unambiguous classifications [17]. Instead, it has been shown that the perceptual structure of object attributes depends on external factors like the context and the focus of attention [18, 50, 56].

The perceptual structure of object attributes does not even imply whether the concurrent manipulation of these attributes is efficient or not. The human locomotor system applies other constraints than the perceived geometry of external objects and environments. More specifically, our motor control seems to use varying reference systems to simplify motion planning, many of which are related to skeletal joints and external forces like gravity [5]. For accurate placement, however, we must consider external reference systems. Consequently, a tendency for concurrent or integral manipulation of many degrees of freedom can be observed during the ballistic phase of movements [26, 46, 66, 67]. Towards the end of an aimed movement, when accuracy becomes key, advantages of DOF separation have been observed instead [2, 48, 62].

Rosenbaum et al. proposed internal representations of posture as a basic building block of human motor planning, because these allow simpler internal representations than trajectories [53, p.178]. Postures can be specified in terms of equilibrium points for the muscles, i.e., "a set of muscle lengths for which muscle tensions balance out". They argue, that "when an equilibrium point is specified and the starting point is known, the trajectory to the equilibrium point comes for free, making detailed planning of the trajectory unnecessary." They also refer to the "end-state comfort effect" which describes a behavioral tendency to grasp objects in an uncomfortable posture in order to achieve a comfortable posture at the end of the manipulation process (Fig. 2). Simultaneous rotation and translation of objects may be inefficient in terms of motion trajectories in world coordinates, but it may be the result of efficient limb coarticulation (see also [52, p.22]).

Simultaneous object rotation and translation can be observed in quotidian activities, although both types of movement are clearly not perceptually integral attributes of object manipulation tasks. Instead, this motor behavior seems to be rooted in the structure of the human

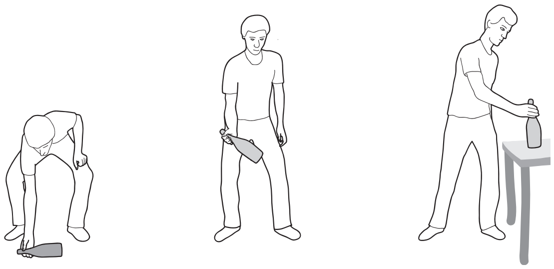


Fig. 2: A bottle lying on the floor, will generally be picked up with an underhand grip to facilitate a comfortable hand posture at the end of the movement.

motor apparatus and it is not necessarily more efficient. Concrete performance benefits have not been demonstrated so far. More constrained object manipulation can be beneficial or even necessary to achieve higher placement accuracy, but, this also leads to more indirect mappings. If the task requires adjustments on multiple degrees of freedom, the resulting segmentation into subtasks is an overhead, that seems unnecessary during coarse target approximation. Smart interaction techniques like snap dragging [6], implicit constraints [60], and velocity-dependent transfer functions [15] can support both, rapid target approximation and placement accuracy. Simultaneous rotation and translation was not always considered in the design of these techniques, but at least, they can be extended in this sense. However, the design, application and parametrization of such techniques is not trivial. More research data on feasible placement accuracy with unconstrained 3D manipulation techniques, could help to better understand, when switching to more constrained interaction modes will be most useful.

3 A FORMAL STUDY ON 3D OBJECT MANIPULATION

The results of prior studies on user performance in 3D manipulation tasks are inconclusive and partially even contradictory. Reference performance models for 3D pointing tasks can be found, but they fail to explain the long task completion times in earlier experiments on 3D object rotation and placement. Reasonable performance benchmarks for 3D object manipulation with six DoF could not be obtained from related work. Consequently, we devised an experiment to establish such reference measures and test the following hypotheses:

H1: Performance in 3D object manipulation tasks follows Fitts's law.

If cognitive factors like mental rotation are controlled and the limits of motor accuracy are avoided, user performance in 3D object manipulation tasks should follow Fitts's law, independent of the type of involved movements (rotation, translation, or both).

H2: Simultaneous rotation and translation can be advantageous.

In real-world settings we typically perform both movements simultaneously. This does not necessarily reduce the movement effort of each subtask, but it avoids overheads of task separation that are related to spatial cognition and motor control.

H3: Oblique target orientations are harder to reach.

3D aimed movements are affected by constraints of our motor apparatus. During reciprocal 3D rotations, people will assume a comfortable hand posture for the more common object orientation, which makes reaching the other one more challenging.

3.1 Experimental Task

We implemented a reciprocal target acquisition task corresponding to Fitts's original experiment from 1954. In this type of studies on motor performance, target distance and target width are typically considered as two independent variables, although Fitts's model suggests that movement time depends on their ratio, not both parameters individually. Guiard argued that, within the limitations of the human motor apparatus, performance in aimed movements should be independent of their scale [21]. He thus suggested to vary only target width while fixing distance at a comfortable level to specify a valid set of task difficulties.

Preliminary experiments with explicit visualizations of rotational and translational target widths, similar to Stoelen and Akin [57], indicated that these can be very confusing and distracting in a docking task with six degrees of freedom. We also found that highlighting the manipulated object or the target visualization when the former was held inside given accuracy thresholds, affected the control strategy of test users. Instead of aimed movements towards a predefined location and orientation, we then observed a tendency to perform small erratic object movements close to the target to trigger the expected highlight. We also noted, that rotational and translational target widths in practical object manipulation tasks are typically not visually specified. Not least, the achievable motor accuracy in mid-air interaction is highly user-dependent. Target distances, instead, are always clearly defined requirements of any 3D manipulation task.

As a consequence, we decided to vary target distance with accuracy set on a user-defined level. The participants of our study were instructed to find their subjective optimum between accuracy and rapidity during training and try to maintain a chosen level of accuracy during the recorded trials. Our post-hoc data analysis revealed only minimal deviations of translational accuracy. For angular accuracy, instead, we observed mean differences of up to 50 % as an effect of target orientation. Each participant's individual deviations were incorporated in the computation of effective target widths for rotation and translation per user and subtask.

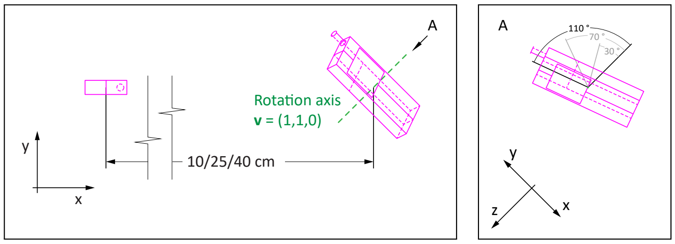
We did not vary the axes of translation and rotation in order to keep the set of test conditions small. Prior research on 3D rotation and translation found significant effects of movement directions. We believe that these could be found in our settings too, but these differences were typically small and not very relevant to answer our main research questions. Therefore, the implemented translation distances required only the reduction of a lateral offset on the x-axis. In case of all tasks involving translation, two alternating target locations were located equidistantly left and right of the screen center. All rotation angles were applied around the vector $\mathbf{v} = (1,1,0)$, a 45 ° diagonal on the screen plane (Fig. 3). During the separate *rotation* and *translation* tasks, the manipulation object could only be rotated (3 DoF) or translated (3 DoF), while it could be moved freely during the *docking* task (6 DoF).

3.2 Conditions

We specified sets of *Rotation* (R), *Translation* (T), and *Docking* (RT) *tasks*. These consisted of *subtasks* with three rotation distances (R = 30 °, 70 °, 110 °) and three translation distances (T = 10 cm, 25 cm, 40 cm). This resulted in 3 *Rotation*, 3 *Translation*, and 9 *Docking tasks*. Within each task set the order of difficulties was fixed from short to long distances. For the *Docking* tasks translation distances were increased before rotation distances. We also distinguished between two *target types* in our reciprocal aimed movement tasks. In case of tasks involving translation, the target could be located *left* or *right*. In case of pure rotation tasks, the targets and the manipulation object remained at the center of the screen, but we distinguish *target types* with *regular* or *oblique* orientation. In case of *3D docking* tasks that combined rotation and translation, the target location on the *left* side always maintained its *regular* orientation in alignment with the display coordinate system, while the target orientation on the *right* side was turned around a diagonal rotation axis on the screen, which resulted in *oblique* orientations (Fig. 3 & Fig. 4).

3.3 Apparatus

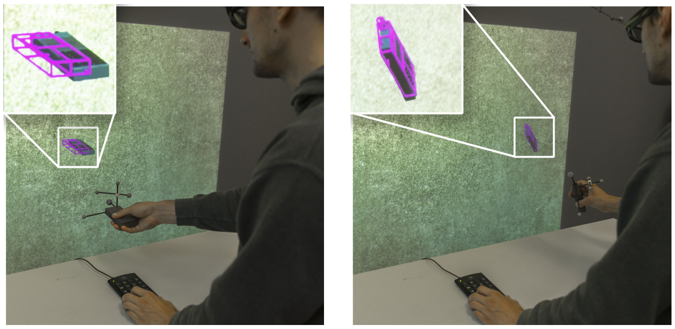
Participants were standing centrally in front of a back projection screen at a distance of about 120 cm. The projected image was 110 cm high and 147 cm wide. Its lower edge was at 70 cm height. Between the screen and the participant was a table of 130 cm width, 50 cm depth, and 100 cm height. The table was standing 50 cm in front of the screen, and thus did not occlude the projected image.



(a) Translation distances along the x-axis and illustration of the rotation axis.

(b) Rotation angles around $\mathbf{v} = (1,1,0)$

Fig. 3: Axonometric visualization of the task conditions: a) view along the depth axis (z), and b) view along the rotation axis **v** as illustrated with the arrow *A* in subfigure (a).



(a) Regular target orientation on the *left* side

(b) Oblique target orientation on the *right* side

Fig. 4: The experimental task required participants to move a 3D object back and forth between two target locations and orientations using a passive 3D tracked prop. The photo was posed to show a monoscopic view for the perspective of the camera. The callouts show a closeup of the virtual object and the docking target with increased contrast.

Participants induced motion input with a passive 3D tracked input device: a wooden block of $11 \cdot 5 \cdot 2$ cm, equipped with optical markers at the front end such that it could be conveniently held and turned between both target orientations with a single hand (Fig. 4). Its position and orientation were captured at 150 Hz with an optical tracking system (ARTRACK5[4]). A numpad was placed on the table for active confirmation of target acquisitions with a keystroke from the second hand. During breaks, the input device could also be laid down at the table.

Visual stimuli were presented with a passive stereo back projection using a pair of CANON XEED WUX10 projectors with circular polarization filters, and lightweight, passive glasses. The glasses were 3D-tracked with the same optical tracking system mentioned above. The projectors did not offer lens shift, hence we used only 89 % overlap between both images. The remaining resolution was 1720 px by 1200 px (~11 px/cm). The experimental task was implemented with a custom 3D application framework. We measured an end-to-end latency from user input to image update of about 100 ms[5].

The manipulation object was represented in the virtual environment with a light blue, bar-type 3D cellphone model. Its size and shape was corresponding to the tangible input device. An antenna, the keypad, and the display were modeled to clearly convey the orientation of the model (Fig. 4). The object isomorphically followed the movements of the input device, but at a fixed offset. We rendered the virtual cellphone model 60 cm closer to the screen and 10 cm higher.

The targets were shown as a purple wireframe representations of the same shape, but 0.25 mm larger in all directions to eliminate z-fighting in case of optimal placement. We are aware that this design implies a certain accuracy, which may have biased our results. Only one of both alternating targets was visible at a time. All targets were displayed with their center position on the screen plane to minimize perceptual conflicts between accommodation and vergence cues. A structured background eliminated the perception of remaining ghosting from the polarization based stereo separation (Fig. 4).

3.4 Procedure

Initially, each participant signed an informed consent form and provided information on age, gender, and handedness. They were required to switch off their cellphone to avoid distraction. Thereafter the experimental setup and the task were explained. A training set of 90 6 DoF-docking trials with nine combinations of 3 translation and 3 rotation distances (slightly different to those of the recorded docking tasks) allowed the participants to become familiar with the experimental task and find their optimal speed-accuracy tradeoff.

After a break of five minutes, we recorded time and accuracy in three task sets: *Rotation* (R), *Translation* (T), and *Docking* (RT). Their order was counterbalanced between three *groups* according to a latin square. Each task set consisted of three or nine subtasks with the above described rotation and/or translation distances (see also Fig. 3). Each subtask was repeated twenty times (ten in each direction). Between each block of 20 repeated subtasks, participants took a break of one to five minutes. The whole experiment lasted between 30 and 45 minutes.

3.5 Participants

We invited twenty-one students of our university via an online blackboard. All had normal or corrected-to-normal vision; ages ranged from 18 to 30 years ($M = 22.8$, $SD = 2.3$); five were female. One participant claimed to be ambidextrous and performed the study with his left hand. All others were right-handed. They received 5 Euro allowance for participation and, to encourage optimal performance, we advertised a restaurant voucher worth 50 Euro for the highest overall throughput. The computation of this measure on the basis of Fitts's law was explained in detail. They were also instructed, to maintain a chosen target acquisition strategy during the recorded trials.

3.6 Results

In our experimental design, both accuracy and time were dependent variables. From the 6300 recorded trials, five were excluded, due to an angular error larger than 15 ° or a translational error larger than 5 cm. The obtained data on rotational and translational inaccuracy as well as task completion times was right skewed (lognormal distributions). We therefore report geometric means. For statistical comparisons between task conditions the data was log-transformed, collapsed, and entered into an analysis of variance. In case of translation and rotation errors, *translation distance*, *rotation distance* and *target type* were taken into account as within-subject factors. Meaningless conditions, i.e. with zero target distance in the relevant dimension, were ignored. For *Movement times*, instead, were compared the effects of 15 *subtasks* and 2 *target types*. *Group* was always tested as a between-subjects factor, but reveled no effect in any of these three tests.

3.6.1 Accuracy

Rotation errors were affected by *rotation distance* ($F_{(2,36)} = 9.86$, $p < .001$, $\eta_p^2 = 0.35$) and *target type* ($F_{(1,18)} = 123.39$, $p < .001$, $\eta_p^2 = 0.87$). The geometric means were ranging from 2.15 ° for 30 ° target distance to 2.42 ° for 110 ° target distance. The effect of *target type* was much larger with 1.87 ° for regularly oriented targets and 2.86 ° for oblique oriented targets. We also found significant interactions of *target type* with *translation distance* ($F_{(3,54)} = 9.71$, $p < .001$, $\eta_p^2 = 0.35$) and *rotation distance* ($F_{(2,36)} = 10.66$, $p < .001$, $\eta_p^2 = 0.37$). A closer look at the data revealed that the effect of *target type* was slightly larger for longer rotational and translational distances. Translational errors were only marginally affected by *target type* ($F_{(1,18)} = 22.95$, $p < .001$, $\eta_p^2 = 0.56$) with geometric means of 3.87 mm for regularly oriented targets on the *left side* and 3.3 mm for oblique oriented targets on the *right side*.

3.6.2 Task Completion Times

We found a significant main effect of *subtask* on *movement times* ($F_{(4.997,89.95)} = 41.55$, $p < .001$, $\eta_p^2 = 0.7$). Mauchly's test indicated a violation of the sphericity assumption ($\tilde{\chi}^2(2) = 191.15$, $p < .001$), hence, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.36$). The geometric means of movement times ranged from 1.68 s for the easiest subtask ($D_T = 10\,\text{cm}, D_R = 0°$) to 2.58 s for the most difficult subtask ($D_T = 40\,\text{cm}, D_R = 110°$)

Movements times were also significantly affected by the *target type* ($F_{(1,18)} = 9.71$, $p < .001$, $\eta_p^2 = 0.35$) with geometrical means of 2.07 s for the regular target orientations on the left side and 2.26 s for the oblique target orientations on the right side. We also obtained a significant interaction effect between *subtask* and *target type* ($F_{(14,252)} = 8.97$, $p < .001$, $\eta_p^2 = 0.33$), which relates to slightly larger effects of *target type* in conditions with longer *rotation distances*.

---

[4] https://ar-tracking.com/products/tracking-systems/arttrack5/

[5] We used a motorized, rotating tracking target and toggled display brightness at a specified orientation, which also triggered a magnetic switch. The visual event was captured with a photodiode on the display.

3.6.3 Linear Regression Analysis

For each participant and *subtask*, we computed the effective target width from absolute error distributions around the requested manipulation target. This approach is conservative. It reflects the largest absolute deviation in any direction. We also took the manipulation goal of the task and not the mean of the actual placements as our measurement reference, since in any practical application, this is the location and/or orientation to be reached. Multiplying the root mean square of absolute distances from the target by 4.133 yields an effective target width around its actual center, that would have been reached in 96 % of all cases (see [39] for comparsion). From these and the given target distances, we computed the corresponding IDs. The average effective target widths were 19.9 mm (SD=6.9 mm) for translations and 12.8 °(SD=3.7 °) for rotations.

The resulting data set consists of 315 combinations of ID and movement time (MT), which was averaged over all participants. (Fig. 1). A regression analysis across the three different types of task yields a high correlation ($R^2 = 0.95$) with an intercept $a = 1.47$ ($SE = 0.056$) and a slope coefficient of $b = 0.17$ ($SE = 0.01$). Computing the regression for each type of task individually, yields slightly different model parameters as presented in Table 1 (see also Fig. 1).

	a (SE)	b (SE)	R^2
3D Rotation	1.48 (0.02)	0.17 (0.01)	0.99
3D Translation	1.13 (0.06)	0.25 (0.01)	0.99
3D Docking	1.71 (0.15)	0.13 (0.02)	0.78
overall	1.47 (0.06)	0.17 (0.01)	0.95

Table 1: Linear regression parameters per task type. Note that the direct comparison of regression coefficients is not very meaningful, since both 3-DoF tasks were only represented by three *subtask* conditions, while data of nine *subtasks* was available in case of *3D Docking*

.

.

3.6.4 Throughput Comparison

Based on the effective index of difficulty, we computed the average throughput (TP) per user and task [12]. The data was then collapsed and entered into an analysis of variance with three manifestations of *task* as within subjects factors and *group* as a between-subjects factor. We found a large and significant main effect for *task* ($F_{(2,36)} = 111.45$, $p < .001$, $\eta_p^2 = 0.89$).

Pairwise comparisons with Bonferroni adjustment of alpha revealed significant differences between each of the three *tasks* (all $p < 0.001$). The mean throughput in the *3D-Docking* task condition was higher ($M_{RT} = 2.64$, SD=0.66) than in the *3D-Translation* task ($M_T = 1.89$, SD=0.45), which was again higher than the mean throughput in the *3D-Rotation* task ($M_R = 1.43$, SD=0.33) (see Fig. 5).

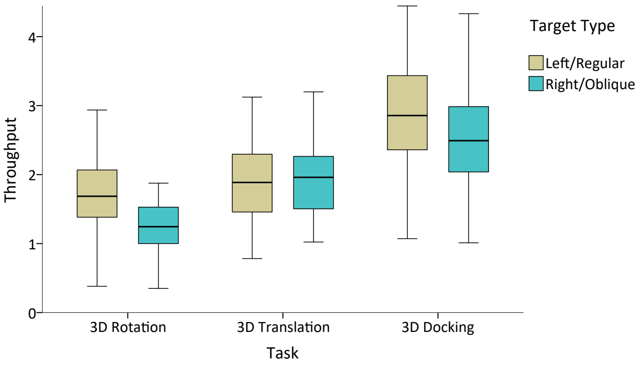


Fig. 5: Boxplots of non-collapsed throughput data per *task* and *target orientation*

3.6.5 Movement Phases

We extracted ballistic movement phases from our log-files using the parsing procedure suggested by Nieuwenhuizen et al. [47]. The heuristic builds on a threshold of 5 % peak velocity to separate movement phases and accumulated distances for their classification. Movement phases that account for at least 25 % of the overall distance are considered ballistic. We found that phases of translational movements were clearly discernible on this basis.The recorded rotational movements, instead, were more erratic. With the suggested threshold of 5 % peak velocity most of the overall movement was classified as ballistic, although, a short moment of much faster rotation can be identified at the beginning of each trial (Fig. 6). Therefore a threshold of 10 % peak velocity was applied for rotations, but the resulting classification is vague.

In pure *translation* tasks, the ballistic phase made up 47.1 % of the overall task duration on average (SD=9.7 %), which resembles the the observations of Nieuwenhuizen et al. [47] in their virtual 3D selection task. In the tasks involving only *rotation* we observed about 38.2 % ballistic movement phase on average (SD=7.8 %). In both cases, target distance had only a marginal effect on the relative phase duration (±1-2 %), hence the absolute duration of ballistic phases was longer for higher target distances.

At the end of the identified ballistic phases, most of the target distance had been covered. On average the remaining distance in the *translation* tasks was 14.1 mm (SD=9.3 mm). As in the studies of Nieuwenhuizen et al., we observed an effect of target distance. The average remaining error at the end of the ballistic phase was 11.2 mm for the 10 cm target (SD=7.5 mm) and 16.3 mm for the 40 cm targets (SD=8.3 mm). At the end of the ballistic phase in the pure *rotation* tasks, the remaining angular error was 8.2 ° on average (SD=2.4 °). It was also dependent on distance with 7.0 ° for the 30 ° target (SD=1.9 °) and 10.6 ° for 110 ° target (SD=2.3 °).

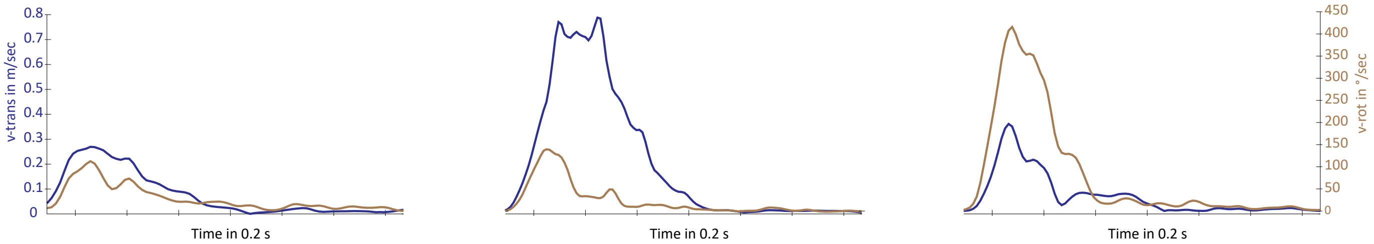
For the *docking* tasks, we analyzed rotation and translation phases separately. For both movements we obtained a smaller ratio of ballistic phases from the overall task duration than in both separate tasks (M_R =32.9 %, SD=9.1 %; M_T =35.4 %, SD=8.5 %). The additional effort of simultaneous rotation and translation seems to be more pronounced during the closed-loop than the ballistic phase. It reflects the earlier observed tendency to separate degrees of freedom towards the end of an aimed movement (when accuracy becomes key) [2, 48, 62].

On average, both ballistic phases were overlapping with about 75 %, which indicates simultaneity of rotation and translation. In the experiments of Wang et al. angular movements were enclosed in translational ones [66]. Our observations are similar for task conditions with short rotation and long translation distances. If it was the other way around, we observed cases in which a short ballistic translation was enclosed by the longer ballistic rotation, and others in which both phases had a similar duration and were largely overlapping (Fig. 6).

3.7 Discussion

The results of our study support hypothesis H1. Average user performance in 3D object manipulation tasks seems to comply with Fitts's law – independent of the type of involved movements. Apparently, 3D object rotation, translation, and the combination of both can be modeled by the same regression coefficients if they are performed under similar conditions and by the same users. Stoelen and Akin's suggestion to simply add the indices of difficulty (ID) for rotation and translation is also supported by our observations. However, our results do not prove compliance with Fitts's law. The predicted scale-independence [21] needs to be tested with conditions that yield the same ID with varying target distances. Moreover, the regression analysis per *task* revealed slightly different parameters, which indicates subtle differences of rotational and translational efforts if performed in combination. Further research is needed to gain a better understanding of these factors.

Our results indicate, that motor performance in isomorphic *3D-Docking* tasks with six DoF can be estimated from performance data on *3D Translation* or *3D Rotation* using a similar experimental apparatus. We thus believe, that the exceptionally long task completion times that were often observed for 3D object manipulation [9, 43, 64, 67, 74] largely stem from other factors than the achievable motor performance, e.g.,



$D_R 30°$, $D_T 10$ cm $D_R 30°$, $D_T 40$ cm $D_R 110°$, $D_T 10$ cm

Fig. 6: Three exemplary sections of velocity-time diagrams visualizing movement phases in different *subtask* conditions. Brown curves show rotation velocity, blue curves show translation velocity.

from cognitive load or excessive accuracy requirements. We suggest to isolate the motor effort from such confounding variables for the evaluation of direct manipulation interfaces. On the basis of our results, diameters of about 20 mm and angular openings of 10-15 ° can be considered comfortable target widths during 3D object manipulation within arm's reach. This indicates rather low accuracy of mid air interactions and highlights the demand for intelligent constraints, adaptive transfer functions, and snapping techniques.

Compared to prior work, the linear regression models for our data have rather shallow slopes ranging from b=0.13 for the *3D-Docking* task to b=0.25 for the *3D-Translation* task. However, this is very similar to the results from Fitts's disc-transfer task (b=0.09 [39], see Fig. 1). It means that increasing the task difficulty by one bit, e.g. twice the distance or half the target width, results in approximately 100 ms longer task duration. Other studies on 3D aimed movements (e.g. [20, 69]) or combinations of rotation and translation [57] observed a stronger influence of ID on movement time (Fig. 1). The shallow slope indicates effective compensation for longer distances through higher velocity. In our case, it may also be related to training, since *subtasks* were presented with increasing difficulty.

We also obtained comparably high intercepts ranging from a=1.13 for the *3D-Docking* task to a=1.71 for the *3D-Translation* task. In other studies on repetetive target acquisitions the computed intercepts were much closer to zero (e.g. [12, 20]). This difference may be partly attributed to system latency [70], but we believe that it is more strongly related to the user-defined placement accuracy in our study. Our participants spent most of the time in the closed loop phase, in which only small improvements could be achieved, but the placement accuracy was visually controlled. The duration of this final verification phase could probably be reduced with suitable system support and better feedback.

Although average performance with the three different movement types in our study could be represented with the same linear regression model, we obtained significantly higher throughput measures for the *3D-Docking* task that combined rotation and translation (RT). Also the throughput of *rotation* and *translation* tasks was found to be significantly different. This may appear contradictory, but it simply reflects, that aimed movements yield a higher throughput for higher IDs if the intercept of the linear regression is positive (see [73]). In other words, it is not the throughput of different movement types that differs, but the throughput of different IDs. In practice, this means that performing aimed movements with a high index of difficulty in one continuous action minimizes the constant overhead of multiple separate ones. Consequently, it was advantageous in our experimental setting to rotate and translate objects simultaneously, which confirms hypothesis H2. Performing both subtasks subsequently takes more time than doing so simultaneously. The sum of average task completion times for separate *rotation* and *translation* is about 40 % longer than that for the combined *docking* task. Switching modes would take additional time.

Last but not least, we observed significantly higher rotation errors and longer task completion times for the oblique oriented target on the *right* side, especially in conditions with large angular target distances, which confirms H3. Simultaneity of rotation and translation, on the other hand, could be observed for both *target orientations*.

4 CONCLUSION AND FUTURE WORK

We searched for reference measures for the cross-validation of studies on direct manipulation interfaces in the literature on aimed movement studies. For 2D pointing, many references can be found with relatively consistent reports of performance based on a throughput analysis using Fitts's law. In the realm of 3D user interfaces and in particular for 3D docking tasks with six DoF, however, results of different studies are highly varying and difficult to relate to each other since a Fitts's law-based analysis was not performed. We demonstrated that Fitts's model can also be applied to such combined tasks consisting of 3D rotation and translation, even if no particular tolerance range is specified and even for rotations with awkward target orientation. Therefore, the effective index of difficulty of most direct manipulation task can be specified, which facilitates the comparison of obtained results to references from the literature.

Our results also offer reference performance measures for isomorphous 3D object manipulation with isotonic devices and minimized cognitive load (reciprocal target acquisition). Apparently, the results compare quite well to performance measures of earlier aimed movement studies, although we observed a large static overhead (Figure Fig. 1). If much lower performance is achieved in a 3D object manipulation task, possible reasons need to be discussed. These may involve overheads imposed by the task (e.g. mental rotation) or the operation of the interface. Zhai showed, for example, that elastic rate control is less effective for 3D docking tasks [74]. Perhaps, ballistic target approximation is not well supported by this type of interface. Rate-controlled motion requires a tight visual feedback loop while users of position control interfaces can take advantage of proprioceptive feedback.

Further studies, with different interfaces and more varying conditions, are necessary to extend our initial collection of reference performance measures. We suggest to use the resulting benchmarks for the cross-validation of results from evaluations of novel interfaces and interaction techniques. This implies to compute effective indices of difficulty for all task conditions in studies on 3D object manipulation with six or less DoF, hence, to adopt this well established research practice from the realm of 2D user interfaces. This will facilitate comparisons among the results of such studies and thereby help to gain a better understanding of the factors that contribute to constant overheads as expressed by the intercept *a* and varying effects of task difficulty as expressed by the slope coefficient *b* in Fitts's law.

We also discussed that the theory of integral and separable perceptual attributes could be misleading for the design of direct manipulation interfaces with multiple degrees of freedom. The corresponding research on human information processing was not concerned with interactive systems, but with the perception and processing of static stimuli [17]. Studies in the field indicate that orientation and position are separable spatial attributes [29]. Human motor control, however, is governed by further factors, in particular, the capabilities and constraints of our body and the affordances of the operated devices [5, 25, 53]. Our study supports this view by clearly demonstrating that simultaneous 3D rotation and translation can be advantageous, despite the perceptual separability of the affected attributes.

ACKNOWLEDGMENTS

We thank the reviewers of this paper for their helpful remarks. We also thank the participants of our studies as well as the members and students of the Virtual Reality and Visualization Research group at Bauhaus-Universität Weimar for their support. Our research has received funding from the German Federal Ministry of Education and Research (BMBF) under grant 03PSIPT5A (project Provenance Analytics).

REFERENCES

[1] R. Arsenault and C. Ware. The importance of stereo and eye-coupled perspective for eye-hand coordination in fish tank vr. *Presence: Teleoper. Virtual Environ.*, 13(5):549–559, Oct. 2004. doi: 10.1162/1054746042545300

[2] R. Balakrishnan, T. Baudel, G. Kurtenbach, and G. Fitzmaurice. The rockin'mouse: Integral 3d manipulation on a plane. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pp. 311–318. ACM, New York, NY, USA, 1997. doi: 10.1145/258549.258778

[3] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger. Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr? In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 585–592, March 2019. doi: 10.1109/VR.2019.8797975

[4] F. Bérard, J. Ip, M. Benovoy, D. El-Shimy, J. R. Blum, and J. R. Cooperstock. Did "minority report" get it wrong? superiority of the mouse over 3d input devices in a 3d placement task. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II*, INTERACT '09, pp. 400–414. Springer-Verlag, Berlin, Heidelberg, Germany, 2009. doi: 10.1007/978-3-642-03658-3_45

[5] A. Berthoz. *The brain's sense of movement: perspectives in cognitive neuroscience*. Harvard University Press, 2000.

[6] E. A. Bier. Snap-dragging in three dimensions. In *Proceedings of the 1990 Symposium on Interactive 3D Graphics*, I3D '90, p. 193–204. Association for Computing Machinery, New York, NY, USA, 1990. doi: 10.1145/91385.91446

[7] G. Casiez, D. Vogel, Q. Pan, and C. Chaillou. Rubberedge: Reducing clutching by combining position and rate control with elastic feedback. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, pp. 129–138. ACM, New York, NY, USA, 2007. doi: 10.1145/1294211.1294234

[8] Y. Cha and R. Myung. Extended fitts' law for 3d pointing tasks using 3d target arrangements. *International Journal of Industrial Ergonomics*, 43(4):350 – 355, 2013. doi: 10.1016/j.ergon.2013.05.005

[9] E. Chapoulie, T. Tsandilas, L. Oehlberg, W. Mackay, and G. Drettakis. Finger-based manipulation in immersive spaces and the real world. In *3D User Interfaces (3DUI), 2015 IEEE Symposium on*, pp. 109–116. IEEE, 2015.

[10] O. Chapuis and P. Dragicevic. Effects of motor scale, visual scale, and quantization on small target acquisition difficulty. *ACM Trans. Comput.-Hum. Interact.*, 18(3):13:1–13:32, Aug. 2011. doi: 10.1145/1993060.1993063

[11] H. Drewes. Only one fitts' law formula please! In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pp. 2813–2822. ACM, New York, NY, USA, 2010. doi: 10.1145/1753846.1753867

[12] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47:381–391, 1954.

[13] P. M. Fitts and J. R. Peterson. Information capacity of discrete motor responses. *Journal of Experimental Psychology*, 67:103–112, 1964.

[14] C. Forlines and R. Balakrishnan. Evaluating tactile feedback and direct vs. indirect stylus input in pointing and crossing selection tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pp. 1563–1572. ACM, New York, NY, USA, 2008. doi: 10.1145/1357054.1357299

[15] S. Frees, G. D. Kessler, and E. Kay. Prism interaction for enhancing control in immersive virtual environments. *ACM Trans. Comput.-Hum. Interact.*, 14(1), May 2007. doi: 10.1145/1229855.1229857

[16] B. Fröhlich, H. Tramberend, A. Beers, M. Agrawala, and D. Baraff. Physically-based manipulation on the responsive workbench. In *Proceedings of the IEEE Virtual Reality 2000 Conference*, VR '00, pp. 5–. IEEE Computer Society, Washington, DC, USA, 2000.

[17] W. R. Garner. *The processing of information and structure*. Lawrence Erlbaum, Oxford, 1974.

[18] W. R. Garner and G. L. Felfoldy. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1(3):225–241, 1970.

[19] E. D. Graham and C. L. MacKenzie. Physical versus virtual pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pp. 292–299. ACM, New York, NY, USA, 1996. doi: 10.1145/238386.238532

[20] T. Grossman and R. Balakrishnan. Pointing at trivariate targets in 3d environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pp. 447–454. ACM, New York, NY, USA, 2004. doi: 10.1145/985692.985749

[21] Y. Guiard. The problem of consistency in the design of fitts' law experiments: Consider either target distance and width or movement form and scale. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pp. 1809–1818. ACM, New York, NY, USA, 2009. doi: 10.1145/1518701.1518980

[22] Y. Guiard and M. Beaudouin-Lafon. Target acquisition in multiscale electronic worlds. *International Journal of Human-Computer Studies*, 61(6):875 – 905, 2004. Fitts' law 50 years later: applications and contributions from human-computer interaction. doi: 10.1016/j.ijhcs.2004.09.005

[23] Y. Guiard, M. Beaudouin-Lafon, and D. Mottet. Navigation as multiscale pointing: Extending fitts' model to very high precision tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pp. 450–457. ACM, New York, NY, USA, 1999. doi: 10.1145/302979.303128

[24] R. Gupta, D. Whitney, and D. Zeltzer. Prototyping and design for assembly analysis using multimodal virtual environments. *Computer-Aided Design*, 29(8):585–597, 1997.

[25] O. Herbort. Optimal versus heuristic planning of object manipulations: A review and a computational model of the continuous end-state comfort effect. *New Ideas in Psychology*, 31(3):291–301, 2013.

[26] M. Herrlich, B. Walther-Franks, and R. Malaka. Integrated rotation and translation for 3d manipulation on multi-touch interactive surfaces. In *Proceedings of the 11th International Conference on Smart Graphics*, SG'11, pp. 146–154. Springer-Verlag, Berlin, Heidelberg, 2011.

[27] K. Hinckley, R. Pausch, D. Proffitt, J. Patten, and N. Kassell. Cooperative bimanual action. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 27–34. ACM, New York, NY, USA, 1997. doi: 10.1145/258549.258571

[28] E. R. Hoffmann, C. G. Drury, and C. J. Romanowski. Performance in one-, two- and three-dimensional terminal aiming tasks. *Ergonomics*, 54(12):1175–1185, 2011. PMID: 22103725. doi: 10.1080/00140139.2011.614356

[29] S. Imai and W. Garner. Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, 69(6):596, 1965.

[30] R. Jacob, R. Silbert, M. Preston, and D. M. FARLANE. Integrality and Separability of Input Devices. In *Proc. of CHI '94*, pp. 3–26, 1994.

[31] M. KHEMIR, U. Zladivar-Colado, and S. GARBAYA. Modeling Human Movement for 3D Manipulation Task. In *International Workshop IHM*, pp. 1–8. Sousse, Tunisia, June 2012.

[32] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, 68(10):603–615, 2010.

[33] A. Kunert, A. Kulik, A. Huckauf, and B. Fröhlich. A comparison of tracking- and controller-based input for complex bimanual interaction in virtual environments. In *Proceedings of the 13th Eurographics Conference on Virtual Environments*, EGVE'07, pp. 43–52. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 2007. doi: 10.2312/EGVE/IPT_EGVE2007/043-052

[34] J. J. LaViola Jr, A. S. Forsberg, J. Huffman, and A. Bragdon. The influence of head tracking and stereo on user performance with non-isomorphic 3d rotation. In *IPT/EGVE*, pp. 111–118, 2008.

[35] L. Liu, R. van Liere, C. Nieuwenhuizen, and J.-B. Martens. Comparing aimed movements in the real world and in virtual reality. In *2009 IEEE Virtual Reality Conference*, pp. 219–222. IEEE, 2009.

[36] P. Lubos, G. Bruder, and F. Steinicke. Analysis of direct selection in head-mounted display environments. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 11–18, March 2014. doi: 10.1109/3DUI.2014.6798834

[37] I. S. MacKenzie and P. Isokoski. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors*

in Computing Systems, CHI '08, pp. 1633–1636. ACM, New York, NY, USA, 2008. doi: 10.1145/1357054.1357308

[38] I. S. MacKenzie, A. Sellen, and W. A. S. Buxton. A comparison of input devices in element pointing and dragging tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pp. 161–166. ACM, New York, NY, USA, 1991. doi: 10.1145/108844.108868

[39] I. S. MacKenzie, R. W. Soukoreff, and C. Pal. A two-ball mouse affords three degrees of freedom. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '97, pp. 303–304. ACM, New York, NY, USA, 1997. doi: 10.1145/1120212.1120405

[40] S. MacKenzie, I. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact.*, 7(1):91–139, Mar. 1992. doi: 10.1207/s15327051hci0701_3

[41] A. Martinet, G. Casiez, and L. Grisoni. Integrality and separability of multitouch interaction techniques in 3d manipulation tasks. *Visualization and Computer Graphics, IEEE Transactions on*, 18(3):369–380, March 2012. doi: 10.1109/TVCG.2011.129

[42] M. R. Masliah and P. Milgram. Measuring the allocation of control in a 6 degree-of-freedom docking experiment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pp. 25–32. ACM, New York, NY, USA, 2000. doi: 10.1145/332040.332403

[43] D. Mendes, F. Relvas, A. Ferreira, and J. Jorge. The benefits of dof separation in mid-air 3d object manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pp. 261–268. ACM, 2016.

[44] D. E. Meyer, R. A. Abrams, S. Kornblum, C. E. Wright, and J. E. K. Smith. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological Review*, 95:340–370, 1988.

[45] A. Murata and H. Iwase. Extending fitts' law to a three-dimensional pointing task. *Human Movement Science*, 20:791–805, 2001.

[46] M. A. Nacenta, P. Baudisch, H. Benko, and A. Wilson. Separability of spatial manipulations in multi-touch interfaces. In *Proceedings of Graphics Interface 2009*, GI '09, pp. 175–182. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 2009.

[47] K. Nieuwenhuizen, J.-B. Martens, L. Liu, and R. van Liere. Insights from dividing 3d goal-directed movements into meaningful phases. *Computer Graphics and Applications, IEEE*, 29(6):44–53, Nov 2009. doi: 10.1109/MCG.2009.121

[48] N. OSAWA and X. REN. A study on approximate and fine adjustments by hand motion in an immersive environment(human-interface basics,¡special issue¿ game programming). *IPSJ Journal*, 48(11):3568–3576, nov 2007.

[49] L. M. Parsons. Inability to reason about an object's orientation using an axis and angle of rotation, 1995.

[50] B. C. Potts, R. D. Melara, and L. E. Marks. Circle size and diameter tilt: A new look at integrality and separability. *Perception & psychophysics*, 60(1):101–112, 1998.

[51] I. Poupyrev, S. Weghorst, and S. Fels. Non-isomorphic 3d rotational techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pp. 540–547. ACM, New York, NY, USA, 2000. doi: 10.1145/332040.332497

[52] D. A. Rosenbaum. Chapter 2 - core problems. In D. A. Rosenbaum, ed., *Human Motor Control (Second Edition)*, pp. 11 – 41. Academic Press, San Diego, second edition ed., 2010. doi: 10.1016/B978-0-12-374226-1.00002-4

[53] D. A. Rosenbaum, R. G. Meulenbroek, and J. Vaughan. 9 - three approaches to the degrees of freedom problem in reaching. In A. M. W. H. R. Flanagan, ed., *Hand and Brain*, pp. 169 – 185. Academic Press, San Diego, 1996. doi: 10.1016/B978-012759440-8/50013-X

[54] F. Sasangohar, S. MacKenzie, I., and D. Scott, Stacey. Evaluation of mouse and touch input for a tabletop display using fitts' reciprocal tapping task. In *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society, HFES 2009*, HFES 2009, pp. 839–843. Human Factors and Ergonomics Society, Santa Monica, CA, USA, 2009.

[55] U. Schultheis, J. Jerald, F. Toledo, A. Yoganandan, and P. Mlyniec. Comparison of a two-handed interface to a wand interface and a mouse interface for fundamental 3d tasks. In *2012 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 117–124, March 2012. doi: 10.1109/3DUI.2012.6184195

[56] B. E. Shepp. On perceiving objects: Holistic versus featural properties. In B. Shepp and S. Ballesteros, eds., *Object Perception: Structure and Process*. Lawrence Erlbaum Associates, Inc, 1989.

[57] M. F. Stoelen and D. L. Akin. Assessment of fitts' law for quantifying combined rotational and translational movements. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(1):63–77, 2010. doi: 10.1177/0018720810366560

[58] W. Stuerzlinger and C. A. Wingrave. *The Value of Constraints for 3D User Interfaces*, pp. 203–223. Springer Vienna, Vienna, 2011. doi: 10.1007/978-3-211-99178-7_11

[59] R. Teather and W. Stuerzlinger. Pointing at 3d targets in a stereo head-tracked virtual environment. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on*, pp. 87–94, March 2011. doi: 10.1109/3DUI.2011.5759222

[60] R. J. Teather and W. Stuerzlinger. Guidelines for 3d positioning techniques. In *Proceedings of the 2007 Conference on Future Play*, Future Play '07, pp. 61–68. ACM, New York, NY, USA, 2007. doi: 10.1145/1328202.1328214

[61] J. J. Tran, S. Trewin, C. Swart, B. E. John, and J. C. Thomas. Exploring pinch and spread gestures on mobile devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, pp. 151–160. ACM, New York, NY, USA, 2013. doi: 10.1145/2493190.2493221

[62] M. Veit, A. Capobianco, and D. Bechmann. Dynamic decomposition and integration of degrees of freedom for 3-d positioning. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, pp. 131–134. ACM, New York, NY, USA, 2010. doi: 10.1145/1889863.1889891

[63] M. Veit, A. Capobianco, and D. Bechmann. An experimental analysis of the impact of touch screen interaction techniques for 3-d positioning tasks. In *Virtual Reality Conference (VR), 2011 IEEE*, pp. 75–82. IEEE, 2011.

[64] V. Vuibert, W. Stuerzlinger, and J. R. Cooperstock. Evaluation of docking task performance using mid-air interaction techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, SUI '15, pp. 44–52. ACM, New York, NY, USA, 2015. doi: 10.1145/2788940.2788950

[65] Y. Wang, C. L. MacKenzie, and V. A. Summers. Object manipulation in virtual environments: Human bias, consistency and individual differences. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '97, pp. 349–350. ACM, New York, NY, USA, 1997. doi: 10.1145/1120212.1120428

[66] Y. Wang, C. L. MacKenzie, V. A. Summers, and K. S. Booth. The structure of object transportation and orientation in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, pp. 312–319. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1998. doi: 10.1145/274644.274688

[67] C. Ware. Using hand position for virtual object placement. *Vis. Comput.*, 6(5):245–253, Nov. 1990. doi: 10.1007/BF01900747

[68] C. Ware and R. Arsenault. Frames of reference in virtual object rotation. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, APGV '04, pp. 135–141. ACM, New York, NY, USA, 2004. doi: 10.1145/1012551.1012576

[69] C. Ware and R. Balakrishnan. Reaching for objects in vr displays: Lag and frame rate. *ACM Trans. Comput.-Hum. Interact.*, 1(4):331–356, Dec. 1994. doi: 10.1145/198425.198426

[70] C. Ware and J. Rose. Rotating virtual objects with real handles. *ACM Trans. Comput.-Hum. Interact.*, 6(2):162–180, June 1999. doi: 10.1145/319091.319102

[71] J. O. Wobbrock, E. Cutrell, S. Harada, and I. S. MacKenzie. An error model for pointing based on fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pp. 1613–1622. ACM, New York, NY, USA, 2008. doi: 10.1145/1357054.1357306

[72] J. O. Wobbrock, A. Jansen, and K. Shinohara. Modeling and predicting pointing errors in two dimensions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 1653–1656. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1979183

[73] S. Zhai. Characterizing computer input with fitts' law parameters: The information and non-information aspects of pointing. *Int. J. Hum.-Comput. Stud.*, 61(6):791–809, Dec. 2004. doi: 10.1016/j.ijhcs.2004.09.006

[74] S. Zhai and P. Milgram. Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, pp. 320–327. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 1998. doi: 10.1145/274644.274689

[75] S. Zhai, P. Milgram, and W. Buxton. The influence of muscle groups on performance of multiple degree-of-freedom input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pp. 308–315. ACM, New York, NY, USA, 1996. doi: 10.1145/238386.238534