Metrics for 3D Object Pointing and Manipulation in Virtual Reality

Metrics for 3D Object Pointing and Manipulation in Virtual Reality

Eleftherios Triantafyllidis, Wenbin Hu, Christopher McGreavy and Zhibin Li
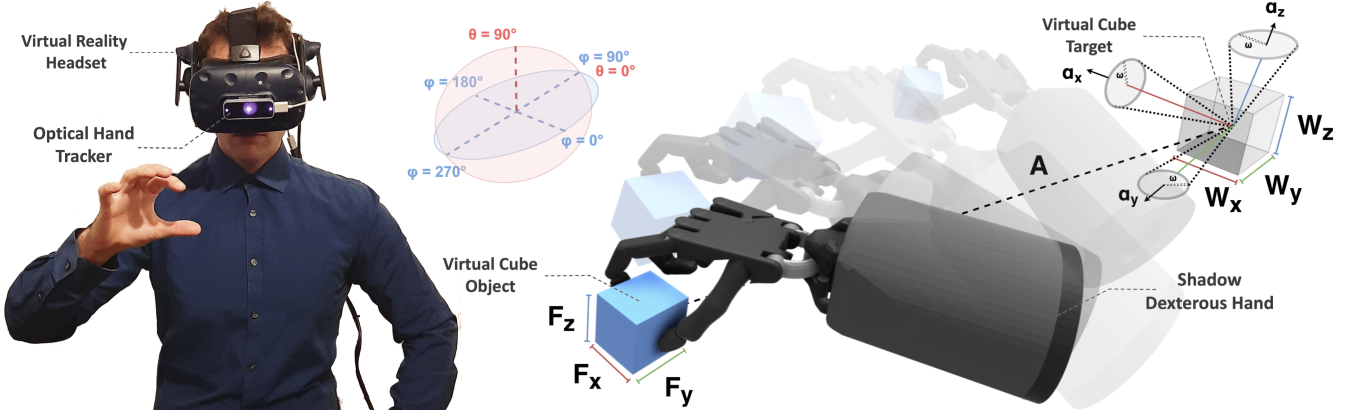
Fig. 1: An operator interacts with objects in full 3D virtual reality with all task-related spatial variables.

Abstract

Assessing the performance of human movements during teleoperation and virtual reality is a challenging problem, particularly in 3D space due to complex spatial settings. Despite the presence of a multitude of metrics, a compelling standardized 3D metric is yet missing, aggravating inter-study comparability between different studies. Hence, evaluating human performance in virtual environments is a long-standing research goal, and a performance metric that combines two or more metrics under one formulation remains largely unexplored, particularly in higher dimensions. The absence of such a metric is primarily attributed to the discrepancies between pointing and manipulation, the complex spatial variables in 3D, and the combination of translational and rotational movements altogether. In this work, four experiments were designed and conducted with progressively higher spatial complexity to study and compare existing metrics thoroughly. The research goal was to quantify the difficulty of these 3D tasks and model human performance sufficiently in full 3D peripersonal space. Consequently, a new model extension has been proposed and its applicability has been validated across all the experimental results, showing improved modelling and representation of human performance in combined movements of 3D object pointing and manipulation tasks than existing work. Lastly, the implications on 3D interaction, teleoperation and object task design in virtual reality are discussed.

I. INTRODUCTION

With the recent advances in networking and Mixed Reality (MR) technologies and in conjunction with the increasingly more immersive applications of telepresence and teleoperation, the need to measure and model human performance in 3D space has increased exponentially [1], [2]. However, a compelling standardized metric for 3D object selection, such as those seen in Virtual Environments (VEs) and teleoperation, does not yet exist. The absence of a standardized metric severely limits inter-study comparability and most importantly transferability between the results of different studies, due to the multitude of different metrics researchers can use [3]. Consequently, progress towards the endeavour of a standardized formulation is still "scattered" and often disregards important aspects that would solidify a concrete and established metric [4].

To propose a higher dimensional metric for assessing human performance, we investigate Paul Fitts' original predictive model, short for Fitts' law [5], [6]. Proposed in 1954, the law has been extensively used in human-computer interaction (HCI) and ergonomics research and still represents the gold standard as a performance metric [2], [7]. This is attributed to the advantage of using Fitts' law to measure human performance in a time-based approach based on spatial data, effectively combining both time and spatial based metrics under one formulation. Fitts' law was originally formulated for 1D translational movements [5], but has also been extended to 2D tasks [7], [8] with its applicability highlighted in rotational tasks as well [9], [10], [11]. Recently, Fitts' formulation was also extended to some extent in 3D space, limited to translational tasks, with numerous reformulations [1], [12], [13]. Yet, these either disregarded important spatial aspects or limited their findings to very specific settings. This aggravates inter-study comparisons particularly due to variations in tested motor tasks.

More specifically, previously proposed 3D metrics extending Fitts' law, have either disregarded combining translational and rotational tasks under one setting [1], [12], [13], assessing directional [10], [11] or inclination variations [1], [10], [11], [13] in one exhaustive study. All of the aforementioned appear to have significant effects on human performance [2], [4]. More importantly, to date, most studies only focused on pointing tasks and by extent, there is a severely limited focus on manipulation tasks, e.g. with the incorporation of grasping and physical properties such as gravity and friction forces.

The authors are with the School of Informatics, The University of Edinburgh, Edinburgh, United Kingdom.
{eleftherios.triantafyllidis,wenbin.hu,c.mcgreavy,zhibin.li}@ed.ac.uk

Consequently, we studied the aforementioned factors and spatial complexities seen in full 3D space to investigate their contribution towards a human performance model. More specifically, we performed an exhaustive user study in Peripersonal Space (PPS) i.e. close to the user's body, in a virtual simulation environment using a VR headset and optical hand tracking. The virtual environment consisted of simulated teleoperation tasks with high-fidelity physics leveraging the anthropomorphic dexterous robotic hand. By collectively combining and adding 3D spatial variables in four progressively more complex experiments, we derived our own model extension and compared the proposed metric with the state of the art while also verifying their applicability at each stage.

From our results, to the best of our knowledge, we present a first of its kind 3D human performance metric based on Fitts' law, which extends beyond current work by modelling full 3D space better than existing formulations. More specifically, our metric is able to capture 3D human motion entailing combined translational and rotational movements, with varying degrees of directions and inclinations in both object pointing and manipulation under a single formulation. This metric can be used to assess human performance by modelling the complex motions, Degrees of Freedoms (DoFs) and dimensions associated with VEs entailing Virtual Reality (VR) [1], [14] as well as teleoperation [3], [15], [16]. Consequently, the effects of different user interfaces, devices and robotic systems on user performance can be modelled and assessed with such a metric with the added advantage of also combining time and spatial based metrics under one model.

The contributions of our work are summarized as follows:

- We propose a new higher-dimensional metric to assess human motor performance in full 3D space;
- An intuitive motion re-targeting of hands in high-fidelity physics to allow generalization of our method towards virtual and realistic object interaction and manipulation;
- A thorough comparison of our proposed metric with others in the literature under different experimental settings and their validity towards higher dimensions;
- Study of quantities and variables to better explain the 3D spatial relationship of objects and recommendations for the design of pointing and manipulation tasks in VR and teleoperation.

In the remainder of the paper, we first introduce the most widely used extension methods based on Fitts' law and those to be compared with our metric. We then describe our methodology, apparatus and the design of the experiments which progressively add more complex spatial variables at each stage. Finally, we derive our proposed metric and discuss and compare our results with the extensions in the related work. A video presentation of this work is also available online[1].

II. RELATED WORK

In this section, we introduce Fitts' original law for translational tasks and the current state of the art of its 2D and 3D extensions, as well as the importance of including rotation into a formulation. Ultimately, we investigate the incorporation of both translational and rotational movements under one combined model. In Table I we summarize all model extensions.

A. *Fitts' Original Formulation*

Fitts' formulation has been extensively used in motion prediction in HCI and ergonomics research [5], [6]. The formulation predicts the *Movement Time* (MT) of how long it takes for users to point to a target on a screen. It is formulated as:

$$MT = a + b \cdot ID, \text{where } ID = log_2\left(\frac{2A}{W}\right).$$

(1)

Here, A represents the distance between the object and the target. W represents the width of the target area. The logarithmic term ID, represents the *Index of Difficulty* of the task, measured in bits per second. The resultant MT is measured in seconds. The constants a and b represent the y-intercept and slope respectively and are derived via regression analysis.

B. *Importance and Limitations of Fitts' Law*

While extensively used, the original formulation shown in Eq. 1 suffers in terms of simplicity when full 3D space is considered. More specifically, the formulation is limited to four-key areas, namely, (i) lower-dimensional 1D and 2D space, (ii) lower DoFs entailing only translational tasks without the combination of rotational movements, (iii) pointing tasks without the addition of physical properties such as gravity or friction e.g. manipulation of objects and (iv) single line movements without the use of spatial arrangements e.g. directions and inclinations. During interactions either in VR or teleoperation, all four aspects are a fundamental and inseparable part of human motion in 3D [2], [10], [14].

Nevertheless, the ability of the law to combine both time and spatial based metrics under a single formulation renders the pursuit of extending it to 3D of significant importance. As identified, part of the motivation of this paper stems from supplementing different evaluation metrics with a single metric in 3D space. The derivation of a model extension based on Fitts' law for full 3D space, is expected to increase the overall comparability between the results of different studies attempting to capture 3D human motion [2], [3], particularly attributed to the significant focus on the law in current work.

---

[1]Video presentation of this work including the virtual tasks completed by the participants: `https://youtu.be/MKH2gGJQq2o`.

Hence, relying on multiple types of metrics to assess human performance in either VR or teleoperation, should ideally be avoided, as comparability between the results of such studies is rendered challenging [2], [16]. Earlier works on teleoperation [16] and recent ones with anthropomorphic robotic hands (19 DoFs) leveraging MR technologies [15], unfortunately, do not make use of metrics such as Fitts' law and instead rely on multiple time-/spatial-/behaviour-based metrics. As a result, comparability even between similar work is aggravated, since all compared studies need to have identical evaluation metrics.

C. 2-Dimensional Extensions

While originally developed for 1D tasks, Fitts' predictive model has also been widely applied to 2D pointing tasks [7], [8], [17]. Hoffmann [8] conducted a series of discrete tapping tasks, using participants' fingers as pointing probes with the width of the finger added to the ID of Eq. 1 as:

$$ID = log_2 \left(\frac{2A}{W + F} \right), \tag{2}$$

where F represents the index finger pad size, which can be interpreted as the size of the object to be transported to the target W as a natural extension towards pick and place tasks [18]. Welford [17] proposed another variant of Eq. 1 as:

$$ID = log_2 \left(\frac{A}{W} + 0.5 \right), \tag{3}$$

which has been demonstrated to do well in 2D task settings.

Mackenzie [7] extended Fitts' original law, known as the Shannon equation, yielding a better fit and formulated as:

$$ID = log_2 \left(\frac{A}{W} + 1 \right). \tag{4}$$

The robustness and linear fit of the Shannon model has been demonstrated for both translational [11], [13] and rotational tasks [11]. The resulting IDs of Mackenzie's model are one bit less than with Fitts's formulation.

Contrary to Fitts' model, the Shannon formulation is limited in terms of mathematical expressiveness. Though Mackenzie argued that the addition of the +1 term avoids negative IDs in the original Eq. 1, this is not entirely true. A negative ID in Fitts' model would mean that the cursor or probe is already within the target area. This limits theoretical justification for the purpose of higher model fitting [4]. Nevertheless, adopting Mackenzie's model facilitates the comparability of numerous extensions as this has become the norm for most of the subsequent models hereinafter presented.

D. 3-Dimensional Extensions

While Fitts' law has been applied to some extent towards 3D pointing tasks [12], [13], it does not accurately represent 3D movement [2], [11]. Murata and Iwase [13] introduced an extension of Fitts' law to 3D pointing tasks taking into account directional angles θ between an origin and a target. Their model is based on Eq. 4 and formulated as:

$$ID = log_2 \left(\frac{A}{W} + 1 \right) + c \cdot \sin \theta, \tag{5}$$

with θ being the azimuth angle and effectively added to the ID. Their work found that directional angles θ had an almost sinusoidal relationship with MT [13].

A later study by Cha and Myung [12] extended the above work by adding inclination angles, representing higher dimensions in finger aimed pointing tasks in the spherical coordinate system. Based on Eq. 2, it is formulated as:

$$MT = a + b \cdot \theta_1 + c \cdot \sin \theta_2 + d \cdot log_2 \left(\frac{2A}{W + F} \right), \tag{6}$$

where θ_1 and θ_2 represent the inclination and azimuth angles from the starting point to the target respectively. Directional angles followed again an almost sinusoidal relationship with MT as shown in Murata's and Iwase's work [13]. The constants a, b, c and d are empirically determined through linear regression. However, in the work of [12], they limited their investigation to forward motions covering between $\pm 30°$ and $\pm 60°$ azimuth angles. Recent work, by Barrera Machuca and Stuerzlinger [1] accounted for the above by introducing pointing tasks with the use of 3D displays to point towards targets ranging from azimuth angles of $-90°$ to $90°$ and $0°$ to $180°$. Their work confirmed that left-to-right movements were easier than movements away from or towards the user. While this investigation covered a wider range of azimuth angles, it still disregarded inclination angles as the height of the objects was adjusted to the view height of each participant. However, human motor skills vary significantly with the direction of movements, for example, upward movements appear to be more demanding than downward ones [12]. Hence, it is important to study the effects of both directions and inclinations.

So far in our analysis, no model has investigated rotational variations or even combined rotational with translational movements all in one setting. Additionally, the identified spatial arrangements and factors in this section should ideally be included as well. Our work aims to address these by effectively combining them in one setting.

Performance Models	Model Formulation and Equation		Model Characteristics				
	MT (sec)	ID (bit/sec)	Derived From	Space	Dir.	Inc.	Rot.
Fitts' [5]	$MT = a + b \cdot ID$	$ID = log_2\left(2A/W\right)$	[N/A]	2D	No	No	Yes
Hoffmann's [8]	$MT = a + b \cdot ID$	$ID = log_2\left(2A/W + F\right)$	[5]	2D	No	No	No
Welford's [17]	$MT = a + b \cdot ID$	$ID = log_2\left(A/W + 0.5\right)$	[5]	2D	No	No	No
Shannon's [7]	$MT = a + b \cdot ID$	$ID = log_2\left(A/W + 1\right)$	[5]	2D	No	No	Yes
Murata and Iwase's [13]	$MT = a + b \cdot ID$	$ID = log_2\left(A/W + 1\right) + c \cdot \sin\theta$	[7]	3D	Yes	No	No
Cha and Myung's [12]	$MT = a + b \cdot \theta_1 + c \cdot \sin\theta_2 + d \cdot ID$	$ID = log_2\left(2A/W + F\right)$	[13], [8]	3D	Yes	Yes	No

TABLE I: Summary of the models that are the most widely used to date in existing work based on Fitts' law and those compared with our proposed model. Each model's characteristics are summarized and whether important spatial variables are accounted for in the 3D domain. Dir.: Directions. Inc.: Inclines. Rot.: Rotation.
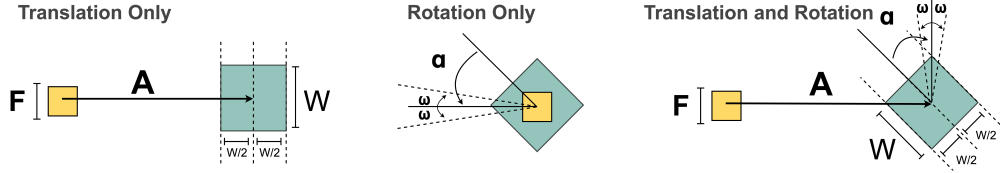

Fig. 2: Illustration of three types of tasks in 2D space: translational tasks, rotational tasks and the combination of both.

E. Combining Translation and Rotation

To this point, we analyzed the most widely used extensions of Fitts' law, limited to purely translational tasks. However, while performing almost any type of manipulation or pointing task, we attempt to match the rotation of the object as well, as to satisfy certain spatial criteria [2], [16], [19]. In addition, to effectively describing general human movement, the simultaneous presence of both translation and rotation needs to be accounted for. Both are an essential and inseparable part when manipulating either real or virtual objects in the 2D and 3D domain [10], [11], [14]. One early study showed that Fitts' law can be adjusted and applied to purely 2D rotational tasks, but did not investigate combined movements [9].

The combination of translational and rotational tasks in 2D space is visually depicted in Figure 2. Stoelen and Akin [11] were the first to "merge" both translational and rotational terms into an extended formulation based on Eq. 4. They both suggested that rotation and translation can be accounted for by effectively adding the sum of indices of difficulty for each one. The implication is that both rotation and translation are independent transformations explaining the spatial relationship of an object. The simplicity of adding both terms in their study into one index of difficulty is supported by Kulik et al. [10]. The latter also found that long task completion times in object manipulation can be attributed to either increased cognitive workload or even to excessive accuracy requirements instructed by the researchers. Yet, the latter finding still requires further modelling of variations of accuracy requirements in order to quantify this relationship.

Nevertheless, in both cases in [10] and [11], combined movements and spatial arrangements were limited to 2D, i.e. following only a straight line. As future work, the authors pointed out that it is important to generalize and extend a metric to three dimensions for modelling virtual reality navigation and teleoperation [11]. Our work addresses these limitations.

III. METHODOLOGY AND STUDY DESIGN

In this section, we introduce our system setup including the apparatus and technologies used. We also present our hand-control approach that allowed participants to interact with the virtual objects simulated by realistic high-fidelity physics.

A. System Overview and Apparatus

Figure 3a illustrates the system overview, including the simulation engine Unity3D, all necessary hardware and software. Two input sensors were used to interact with the simulation. The first sensor was the optical hand tracking device – Leap Motion Hand Controller (LMHC). Equipped with two infrared cameras at 120Hz and with a $135°$ Field of View (FoV), the LMHC allowed to interface between the user's hand movements and the physics engine PhysX in Unity3D. For visual feedback, high-resolution displays were used to limit distance overestimation and degraded longitudinal control, a known issue in VEs [1], [2]. Consequently, the Virtual Reality Head-Mounted Display (VRHMD) HTC Vive Pro was used, with a 2880 x 1600 pixel resolution display and $110°$ FoV at $90Hz$. The photosensors on the VRHMD also represented our second sensor, allowing for head rotations in the VE. Furthermore, ROS# was used to import physics models of robots and objects (Unified Robot Description Format). For all experiments, the LMHC was fixed on the front of the VRHMD.

The physics simulation time-step was set at $1000Hz$ to ensure robust and stable performance with realistic forces and frictions. Finally, to ensure optimal hand tracking performance, lightning conditions were consistent and operational space was limited to about 100cm as the upper maximum reaching bounds from the chest of users. In addition, a low-pass filter with a cutoff frequency of 10Hz was applied to the LMHC to reduce noise during retargeting, ensure continuity and robustness.

(a) Physics simulation environment and hand setup.
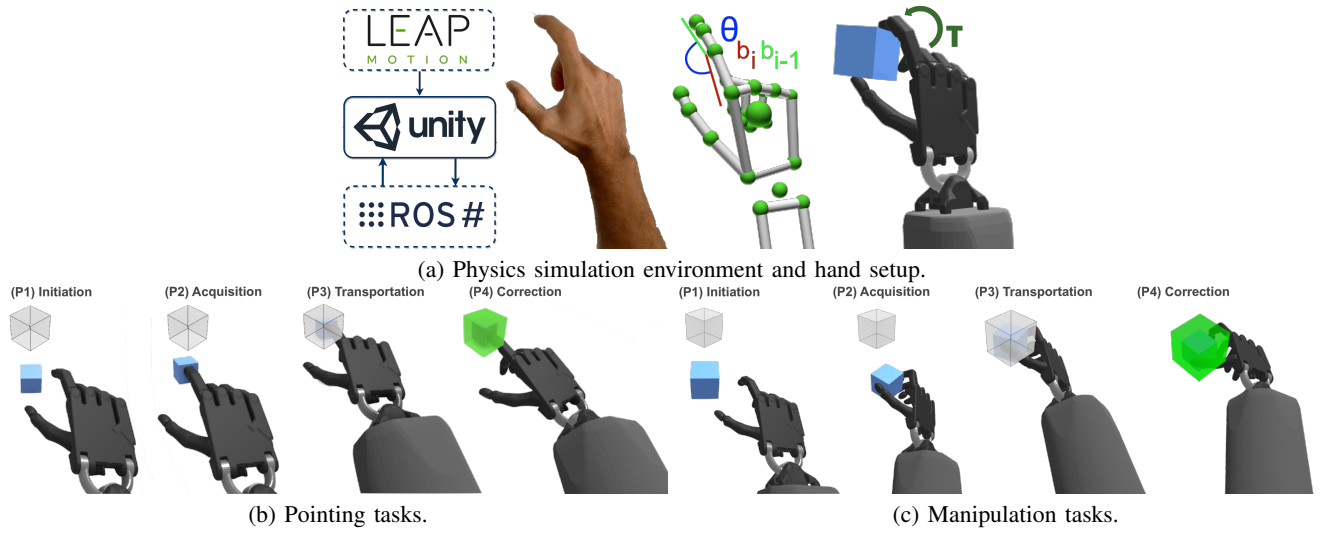

(b) Pointing tasks.

(c) Manipulation tasks.

Fig. 3: The system overview and the two types of interaction in the simulation environment. (a) depicts the simulation environment and the retargeting approach: the illustration of the user's hand, the LMHC visualisation of the joint vectors and the motion retargeting of the robotic hand in Unity3D. (b) and (c) represent the pointing and manipulation tasks respectively, broken down into the four basic phases of interacting with an object. The only difference between (b) and (c) lies in phase P2, where the object is either grasped or attached to the hand depending on the type of interaction.

B. Hand Control and Input Interface

As shown in Figure 3a, the user's hand movements were mapped onto an anthropomorphic Shadow Robotic hand in the simulation. The palm of the simulated hand had 6 DoF and could move freely around the virtual environment in all axes for both translation and rotation. To tele-control the virtual hand, the Cartesian hand movements and joint positions from the participant were mapped onto that in the virtual environment. The re-targeting approach was similar to the work in [14], where the joint positions of the user's fingers were obtained by calculating the angle θ between a joint $\overrightarrow{b_{i-1}}$ and its parent joint $\overrightarrow{b_i}$. The resultant angle θ from the user's finger is then incorporated in a joint PD controller to achieve the desired re-targeting joint motions from the user hand to the simulated robotic one. These are formulated as:

$$\theta = \arccos\left(\frac{\overrightarrow{b_i} \cdot \overrightarrow{b_{i-1}}}{\left\|\overrightarrow{b_i}\right\| \left\|\overrightarrow{b_{i-1}}\right\|}\right),$$

$$\tau_i = K_{P_i} \cdot (\theta_{d_i} - \theta_{c_i}) - K_{D_i} \cdot \dot{\theta}_i,$$

(7)

where θ represents the desired angle for the virtual hand to match. Furthermore, τ are the torques applied to each joint i. θ_{d_i} and θ_{c_i} are the desired angles from the human hand from the LMHC and the current angle of the virtual hand joint in the simulation, respectively. Finally, $\dot{\theta}_i$ is the measured velocity of the virtual hand for computing the damping torque.

Furthermore, a velocity control signal was applied to the palm of the robotic hand based on the real hand's position and orientation. Hence, the 6 DoFs of the robotic hand were controlled by matching its translation and orientation with that of the real one captured by the LMHC. The collision between the robotic hand and the objects in the simulation environment was realised via the built-in PhysX 4.0 engine in Unity3D. We furthermore retained the original joint limits as well as colliders of the virtual hand, as specified in the URDF file of the Shadow robot hand to ensure optimal and realistic actuation.
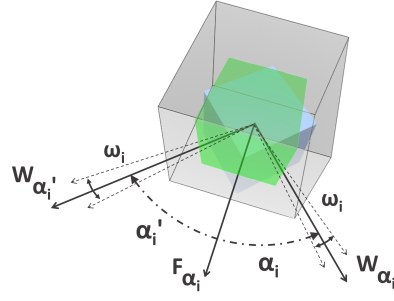
C. Task Design and Spatial Assessment

In each task, participants were asked to move an object from a start to a target location. Contrary to the use of a sphere or the index finger of a participant in most related work [1], [12], [13], the use of a cube allowed us to assess rotational variations in our experiment. Due to its identifiable orientation and as one of the most basic 3D shapes, the cube presented a suitable choice to assess both translational and rotational tasks. Regarding rotation, we instructed participants to match the sides of the cube with that of the cube target, as parallel as possible. While using a cube introduces in essence four "correct" rotations and limits to some extent the range of rotations one can investigate (e.g. to a maximum of 45 degrees), it still represents the dominant and most widely used 3D shape in current work [14]. Our approach is influenced by [10], [11], which included rotational tasks, but were limited to 2D movements only following straight lines without directions and inclinations. The targets were arranged in spherical coordinates with the object at the centre. Figure 2 and Figure 1 illustrate movements in 2D and full 3D space respectively.
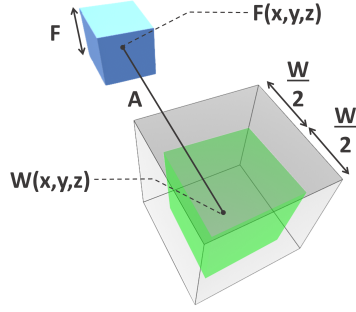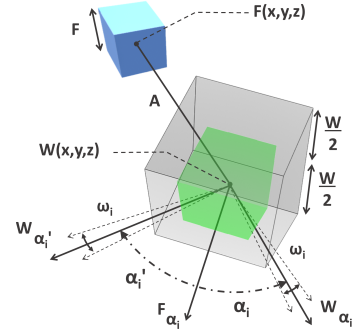
Translational Task   **Rotational Task**   **Combined Task**

Translational Tasks	Rotational Tasks	Combined Tasks
$T = \begin{cases} 1, & A \leqslant \frac{W}{2} \\ 0, & A > \frac{W}{2} \end{cases}$	$R = \begin{cases} 1, & \alpha_i \leqslant \omega_i \\ 0, & \alpha_i > \omega_i \end{cases}, \; i : \mathbb{R}^3$	$C = T \cdot R$

Notations for Translational Tasks: A represents the distance between two 3D points, hence $A = d(W_{x,y,z}, F_{x,y,z}) = \sqrt{\sum_{i=1}^{3}(x_i - y_i)^2}$.

Notations for Rotational Tasks: $\alpha_i = |W_{\alpha_i} - F_{\alpha_i}|$. W_{α_i} and F_{α_i} represent the rotation of the target and the object in all axes (i) respectively. $\alpha_i' = \alpha_i$.

Notations for Combined Tasks: Both translational and rotational requirements need to be met to be classified as success e.g. $T = 1$ and $R = 1$.

Fig. 4: From left to right, the figure represents the progression requirements for Translational, Rotational and Combined tasks. The green cube volume illustrates successful placements. The table represents the task successes (1) or errors (0).

The interactive object manipulated by users was presented as a solid blue cube and the target location as a transparent cubic volume. When intersected and translational and/or rotational requirements met, depending on the task type and difficulty, the transparent target would turn green, indicating the success of the task and progressing to the next, as shown in Figure 3c and Figure 3b. In the cases of pure translational tasks, a 50% overlap with the target was considered a success as with Fitts' original experiment. For rotational tasks, the target rotation, α, needed to be matched within a certain rotation tolerance ω, e.g. the overlap of object-to-target had to be matched in all axes. In Figure 4 we visually depict and describe the mathematical equations that needed to be satisfied in order for each type of task to be classified either as a success or an error.

D. *Pointing vs Manipulation*

In addition to varying all possible spatial variables in 3D object interaction, we also investigated pointing and manipulation tasks, which are the most common types of interactions in both VEs [1] and teleoperated simulation environments [14], [16]. Pointing tasks were investigated as still being the dominant interaction type in 3D user interfaces and it closely resembles Fitts' original experiment. Furthermore, pointing tasks are a natural extension of peg insertion tasks, commonly seen in teleoperation [16]. To cover the most critical limitation in current literature in using Fitts' law for robotics, we also studied manipulation tasks e.g. using grasping, as these are the dominant types of object interaction in teleoperation [2].

For pointing tasks, we used the index pad finger of the virtual hand, to "attach" the cube to the participant's hand, allowing users to move their hand and the cube to the target location. The object would attach or collide during the intersection between the index finger and the object, and it would match the position of the pad index finger but retain its original orientation. On the other hand, for the manipulation tasks, realistic physics and friction forces were simulated, where participants had to grasp and transport the object to the target location by simulated contact forces. Gravity was enabled for manipulation and a table was simulated by a collision plane. Figure 3b and Figure 3c visually depict these two types of object interactions.

IV. ANALYSIS AND PROCEDURE

All results have been tested for significance (95% CI) via the use of a Repeated-Measures ANOVA (RM-ANOVA). A Shapiro-Wilk Test was used to verify the normality of the data prior to the RM-ANOVA. Maulchy's Test of sphericity was used to test sphericity and in cases of violation, a Greenhouse-Geisser Correction (GGC) was used to account for the violation and correct the degrees of freedom assuming $\epsilon < 0.75$. For non-parametric data violating normality, an Aligned-Rank Transform (ART) [20], was used to allow the use of the RM-ANOVA on the ranked data. A step-wise linear regression determined the effect of each variable for conceiving our metric equation and the effect on the predictability of MT. The independent spatial variables were either fitted using the criteria of the probability of $F < .05$ to enter, or $F > .10$ to remove. Finally, we used linear regression analysis (r^2) to analyze and compare our proposed model with existing work. Hereinafter, the significance levels are: *$p < .05$, **$p < .01$, ***$p < .001$, and n.s standing for "not significant".

Variables		Variations Investigated			
		Experiment 1	Experiment 2	Experiment 3	Experiment 4
F	Object Size	(3, 4, 5)	(5)	(4, 5)	(4)
W	Target Width	(5, 7.5, 10, 12.5)	(5, 10)	(5, 10)	(4, 8)
A	Target Separation	(12, 24, 36, 48)	(12, 24)	(0)	(12, 24)
ϕ	Direction Angle	(90)	(0, 90, 180, 270)	(0)	(0, 90)
θ	Inclination Angle	(0)	(15, 30, 45)	(0)	(15, 30)
α	Angular Distance	(0)	(0)	(15, 30, 45)	(30, 45)
ω	Angular Tolerance	(0)	(0)	(2.5, 5, 7.5, 10)	(7.5, 15)
Variations x (Reps.)		48x(5)=240	48x(5)=240	48x(5)=240	64x(4)=256

TABLE II: Parameter settings of all spatial variations.

A. *Participants*

A total number of 20 participants ($N = 20$) were recruited in this study (4 females and 16 males), with ages ranging from 19 to 46 ($\mu = 27.35$, $\sigma = 5.43$). The selection criteria we set during recruitment was that each participant was (i) right-handed, (ii) had healthy hand control with (iii) normal/corrected vision and (iv) was familiar with either video games or VR. Participants that did not meet all of these criteria were excluded from the experiment. Participants were asked to find a balance between minimizing errors and selecting the targets as quickly as they could during target selection and placing.

B. *Pre-Exposure and Approach*

Prior to the commencement of the experiment, participants were briefed, gave formal written consent and their Individual Interpupillary Distance (IPD) was measured for the VRHMD. Furthermore, acclimatization to the simulation environment was allowed for all participants prior to the experiment, via a set of 96 training exercises. These covered both pointing and manipulation task types, e.g. 48 for each type. Furthermore, both translational and rotational movements were included in these exercises, as to cover the entirety of all spatial complexities later investigated. The training exercises were specific to the acclimatization procedure and independent of those in the experiment. We also randomized the order of all experiments for all participants to counterbalance potential acclimatization or task adaption. A grand total of 39,040 trials were recorded.

C. *Experiment Procedure*

To investigate the applicability of our method in full 3D space, we conducted four increasingly more complex experiments as shown in Table II. The breakdown in experiments allowed us to compare existing approaches in increasingly more complex spatial settings. All four experiments included two additional and different types of interactions: pointing and manipulation. Throughout, positional tracking was disabled but head tracking was allowed. Participants were nonetheless asked to retain their body pose without moving their torso. To limit potential tiredness, a resting break of 15 seconds took place for every 15 tasks, where users could rest their hands. To control the duration of all experiments, a maximum number of 15 (pointing) and 20 (manipulation) seconds were given for each task, after which the task ended and was considered an error.

Finally, as manipulation inherits multiple different grasping techniques, generally argued to be six distinctive grasping types according to the Southampton Hand Assessment Procedure (SHAP), we instructed participants to use precision/tip grasping for all manipulation tasks, as shown in Figure 3a and Figure 3c. Error trials were excluded from the analysis. All of the aforementioned settings described in this section remained constant for all experiments.

V. EXPERIMENTS

Here we present all experiments (E1 to E4) with increasing spatial complexity, each has their results interpreted and analyzed for both pointing and manipulation. Table II illustrates all the variations of the investigated variables. Figure 5 visually depicts the relationship between these variables and MT; and Table III summarizes the statistical results and the significance levels. Finally in Table IV and Figure 6, we summarize and compare all models with their respective r^2 values.

A. *Experiment 1: Purely Translational Tasks*

In this experiment, we were primarily interested in approximating Fitts' original experiment, except in 3D. Hence, we investigated the effects of purely translational tasks on the variables of object size F, target width W and target separation A. Movements were along one line only, with no directions or inclinations as to keep the 3D task simple.

1) Design: This study used a 3x4x4 within-subjects design, and the independent variables were: three object sizes ($F = 3$, 4 and 5cm), four target sizes ($W = 5$, 7.5, 10, 12.5cm) and four target separations ($A = 12$, 24, 36, 48cm). The dependent variable was MT. Among 48 types of tasks, each had 5 repetitions for both pointing and manipulation, hence a total of 480 trials. With 20 participants, a total of 9600 trials were recorded.

2) Pointing Results: The data was not normally distributed, as the Shapiro-Wilk Test yielded $p = .029$, and thus an ART was used prior to the RM-ANOVA to allow the analysis on the ranked data. The mean MT for pointing tasks was 1.63 ± 0.40s. Four trials out of 4800 were excluded from the analysis due to errors. Our results showed that an increase in object size F and target width W significantly decreased MT, ($p < .01$) and ($p < .001$) respectively. Compared to F and W, the target separation A had the biggest effect on MT, with an almost linearly increasing correlation ($p < .001$).

	Pointing							Manipulation						
	Sphericity Test			RM-ANOVA Test				Sphericity Test			RM-ANOVA Test			
	χ^2	p	ϵ	F-Test	η_p^2	p	r^2 (%)	χ^2	p	ϵ	F-Test	η_p^2	p	r^2 (%)
					Experiment 1									
F	$\chi^2(2){=}2.468$	n.s	n/a	F(2,30)=9.441	0.386	**	0.5%	$\chi^2(2){=}3.314$	n.s	n/a	F(2,30)=14.468	0.491	***	3.1%
W	$\chi^2(5){=}6.992$	n.s	n/a	F(3,33)=102.522	0.903	***	10.2%	$\chi^2(5){=}5.409$	n.s	n/a	F(3,33)=19.162	0.635	***	5.3%
A	$\chi^2(5){=}8.360$	n.s	n/a	F(3,33)=602.976	0.982	***	80.4%	$\chi^2(5){=}5.968$	n.s	n/a	F(3,33)=197.465	0.947	***	84.7%
					Experiment 2									
W	n/a	n/a	n/a	F(1,23)=287.752	0.925	***	40.0%	n/a	n/a	n/a	F(1,23)=51.431	0.691	***	35.6%
A	n/a	n/a	n/a	F(1,23)=190.993	0.893	***	43.7%	n/a	n/a	n/a	F(1,23)=33.986	0.596	***	23.0%
ϕ	$\chi^2(5){=}12.609$	*	0.675	F(2.025,22.270)=6.918	0.386	*	3.0%	$\chi^2(5){=}3.722$	n.s	n/a	F(3,33)=0.851	0.072	n.s	1.0%
θ	$\chi^2(2){=}4.346$	n.s	n/a	F(2,30)=7.160	0.323	*	1.3%	$\chi^2(2){=}2.210$	n.s	n/a	F(2,30)=5.535	0.270	*	5.4%
					Experiment 3									
F	n/a	n/a	n/a	F(1,23)=1.551	0.063	n.s	0.0%	n/a	n/a	n/a	F(1,23)=1.195	0.049	n.s	0.5%
W	n/a	n/a	n/a	F(1,23)=0.299	0.013	n.s	0.0%	n/a	n/a	n/a	F(1,23)=40.805	0.640	***	6.4%
α	$\chi^2(2){=}0.781$	n.s	n/a	F(2,30)=23.639	0.612	***	1.8%	$\chi^2(2){=}3.589$	n.s	n/a	F(2,30)=12.915	0.463	***	2.9%
ω	$\chi^2(5){=}11.386$	*	0.639	F(1.918,21.093)=184.323	0.944	***	81.8%	$\chi^2(5){=}4.454$	n.s	n/a	F(3,33)=89.481	0.891	***	73.3%
					Experiment 4									
W	n/a	n/a	n/a	F(1,31)<0.001	0.0	n.s	0.0%	n/a	n/a	n/a	F(1,31)=58.815	0.655	***	13.1%
A	n/a	n/a	n/a	F(1,31)=51.296	0.623	***	8.5%	n/a	n/a	n/a	F(1,31)=48.664	0.611	***	19.2%
ϕ	n/a	n/a	n/a	F(1,31)=2.278	0.680	n.s	0.7%	n/a	n/a	n/a	F(1,31)=0.205	0.007	n.s	0.1%
θ	n/a	n/a	n/a	F(1,31)=1.718	0.053	n.s	0.2%	n/a	n/a	n/a	F(1,31)=2.990	0.088	n.s	0.7%
α	n/a	n/a	n/a	F(1,31)=10.548	0.254	**	1.2%	n/a	n/a	n/a	F(1,31)=10.193	0.247	**	4.0%
ω	n/a	n/a	n/a	F(1,31)=394.944	0.927	***	77.3%	n/a	n/a	n/a	F(1,31)=122.753	0.798	***	44.3%

TABLE III: Results of all variables and their influence towards MT, across all experiments. Percentages from r^2 (%) indicate the weight or contribution of each variable towards MT and were obtained via step-wise linear regression.

3) Manipulation Results: The data was normally distributed ($p = .229$) and the mean MT was 2.13 ± 0.45s. In total 44 trials out of 4800, i.e. 0.9% of the data, were excluded from the analysis due to errors. MT significantly decreased as object size F and target width W increased, ($p < .001$) and ($p < .001$) respectively. The target separation A had again the biggest effect, almost linearly increasing with MT ($p < .001$).

4) Remarks: From our analysis, we observed that all included variables had a significant effect on MT for both pointing and manipulation. Fitting the models of Fitts', Shannon's and Welford's using the data, the results showed a high fitting for pointing, but a slightly less fitting for manipulation. Only Hoffmann's model showed a significantly better fitting for manipulation, at an approximately 10% increase from the rest, which is likely due to the incorporation of the object size F in the formulation. Murata & Iwase' and Cha & Myung's were excluded from this analysis as they yielded the same results as their based extensions, since directions and inclinations were not assessed in E1.

B. Experiment 2: Translational Tasks with Directions and Inclinations

E2 is an extension of E1 by adding directions and inclinations, which have significant effects according to [12], [13].

1) Design: The study used a 2x2x4x3 within-subjects design, and the independent variables were: two target sizes (W = 5 and 10cm), two target separations (A = 12 and 24cm), four direction angles ($\phi = 0°$, $90°$, $180°$ and $270°$) and three inclination angles ($\theta = 15°$, $30°$ and $45°$). The dependent variable was MT. A total of 9600 trials were recorded.

2) Pointing Results: The data was normally distributed ($p = .077$) and the mean MT for pointing tasks was 1.47 ± 0.23s. A total of 3 trials out of 4800 were excluded from the analysis due to errors. The target width W significantly decreased MT with higher dimensions ($p < .001$), while high values of target separation A significantly increased MT ($p < .001$). Directional angles ϕ significantly affected MT as well ($p < .05$), with a slight sinusoidal relationship with MT and revealing that front-to-backward movements ($0°$, $180°$) take slightly longer time than left-to-right ones ($90°$, $270°$), in line with [1]. Finally, higher degrees of inclinations θ translated in longer MT, ($p < .05$), matching the findings of [12].

3) Manipulation Results: The data was normally distributed ($p = .083$) and the mean MT for manipulation tasks was 2.11 ± 0.25s. A total of 60 trials out of 4800, 1.25% of the data, were excluded from the analysis due to error. MT significantly decreased as target width W increased ($p < .001$). Consistent with E1, an increase in target separation A significantly increased MT ($p < .001$). Contrary to pointing tasks, directional angles ϕ did not have an apparent effect on MT ($p = .476$). Finally, increased incline angles θ significantly affected MT ($p < .05$) as in [12].

4) Remarks: All variables had significant effects on MT, except directional angles ϕ in manipulation. Contrary to E1, we observed that Fitts', Shannon's and Welford's formulation had a relatively higher fitting both in pointing and manipulation than Hoffmann's model, by approximately 5%. This is confirmed with Cha & Myung's model, which is based on Hoffmann's, having a very similar fitting for both pointing and manipulation. Only Murata & Iwase's directional model presented a marginally higher fitting than the rest in pointing and manipulation tasks.

C. Experiment 3: Purely Rotational Tasks

To this point, only translational tasks were investigated. Yet, from the identified literature and intuition, rotation is an inseparable and fundamental part [2], [10], [11]. Hence, here translation is eliminated as both the object and the target will
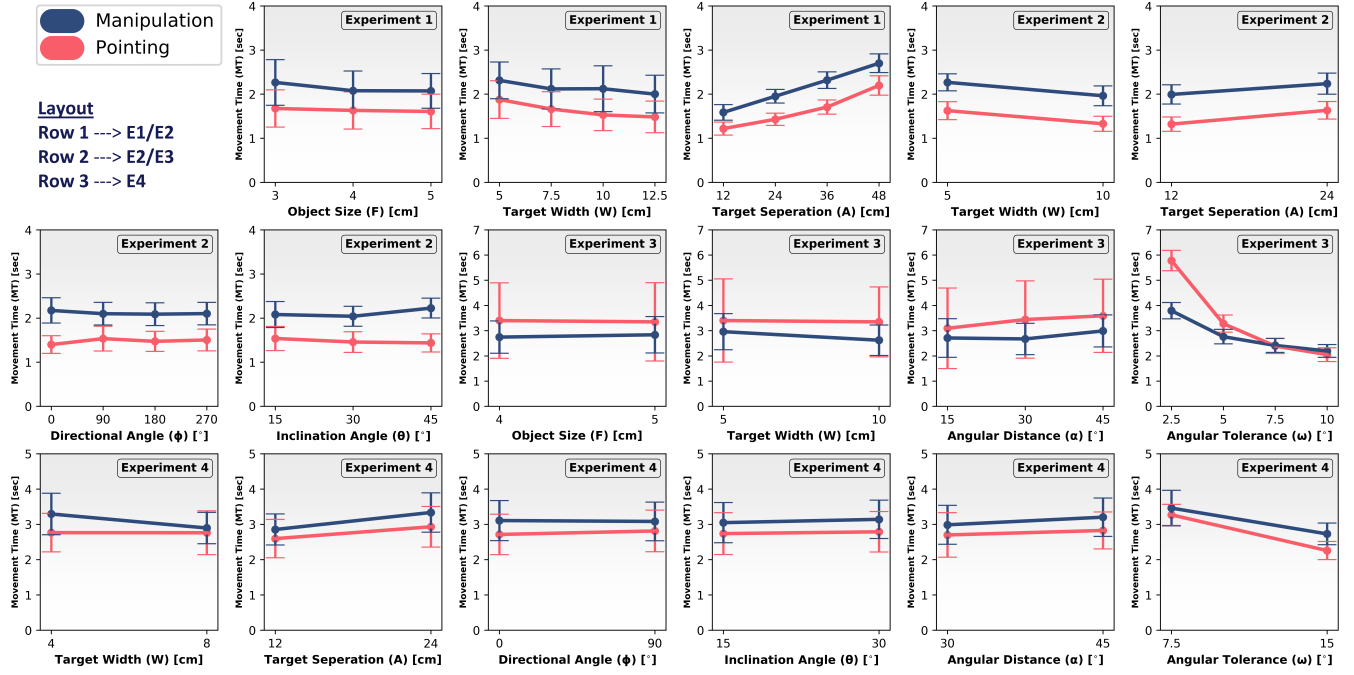
Fig. 5: Results of all tested variables in all experiments and the relationship with MT for both pointing (red) and manipulation (blue). Figure sequence is from left to right and top to bottom. Bars represent the standard deviation.

be overlapping in the centre, with only their rotation differing. Consequently, this task is an almost natural equivalent of Fitts' original experiment, only for rotation.

1) Design: The study used a 2x2x3x4 within-subjects design, in which the independent variables were: two object sizes (F = 4 and 5cm), two target sizes (W = 5 and 10cm), three target rotations (α = 15°, 30° and 45°) and four rotational tolerances (ω = 2.5°, 5°, 7.5° and 10°). The dependent variables was MT. 9600 trials were recorded.

2) Pointing Results: The data deviated from normality ($p < .001$) and an ART was applied, and the mean MT for pointing tasks was 3.37 ± 1.51s. A total of 225 trials out of 4800, 4.6% of the data, were excluded from the analysis due to error. An increase in both object size F and target width W showed a slight decrease towards MT, but not at a significant level, ($p = .226$) and ($p = .590$) respectively. However, we observed a significant correlation between the rotational separation α and MT ($p < .001$). The largest effect was observed with the rotational tolerance ω, showing an almost inverse exponential relationship with MT ($p < .001$). We can infer that the effect of ω was so significant that it had almost the same inverse effect on MT as the equivalent of W and F for translational tasks, though at a higher degree.

3) Manipulation Results: The data violated normality ($p < .01$) and an ART was thus used. The mean MT for manipulation tasks was 2.79 ± 0.67s. A total of 64 trials out of 4800, 1.3% of the data, were excluded due to error. In line with pointing, no significant effect was observed on MT with the object size F ($p = .286$). In contrast to pointing, an increase in target width W did significantly decrease MT ($p < .001$). Increasing rotational separation α also increased MT significantly ($p < .001$). Finally, as with pointing, increasing rotational tolerance ω significantly decreased MT ($p < .001$) in a non-linear fashion. With the exception of target width W, we observed the same relationship of the tested variables on MT for both pointing and manipulation. Rotational tolerance ω predominately affected MT in a purely rotational setting. This was further investigated in the final experiment E4 to increase our understanding of the underlying factors.

4) Remarks: For this experiment, we fit all existing equations to accommodate the rotational nature of this experiment, as detailed in Table IV. By doing so, we are able to conduct and apply these methods to the rotational nature of this experiment, and also investigate any significant effects for comparison. We further extend the work of [11] and cover all known extensions. Overall, regarding variables influencing MT, rotational variables α and ω were significant in both cases, although some inconsistencies were observed between pointing and manipulation. Noticeably, none of the adjusted models fitted this rotational task well, where all formulations were below $r^2 \leq 0.680$ for both pointing and manipulation. Contrary to the translational experiments (E1 & E2), we see that these models did not fit the data very well in rotational tasks.

D. *Experiment 4: Combined Translational and Rotational Tasks with Directions and Inclinations*

For E4, we combined all possible variations from E1, E2 and E3, constituting this experiment a fully combined movement task. We combine both translational and rotational task variations with varying directional and inclination gains to describe a full 3D task. To limit the number of variations, we restricted the directional movements ϕ towards the front and right side
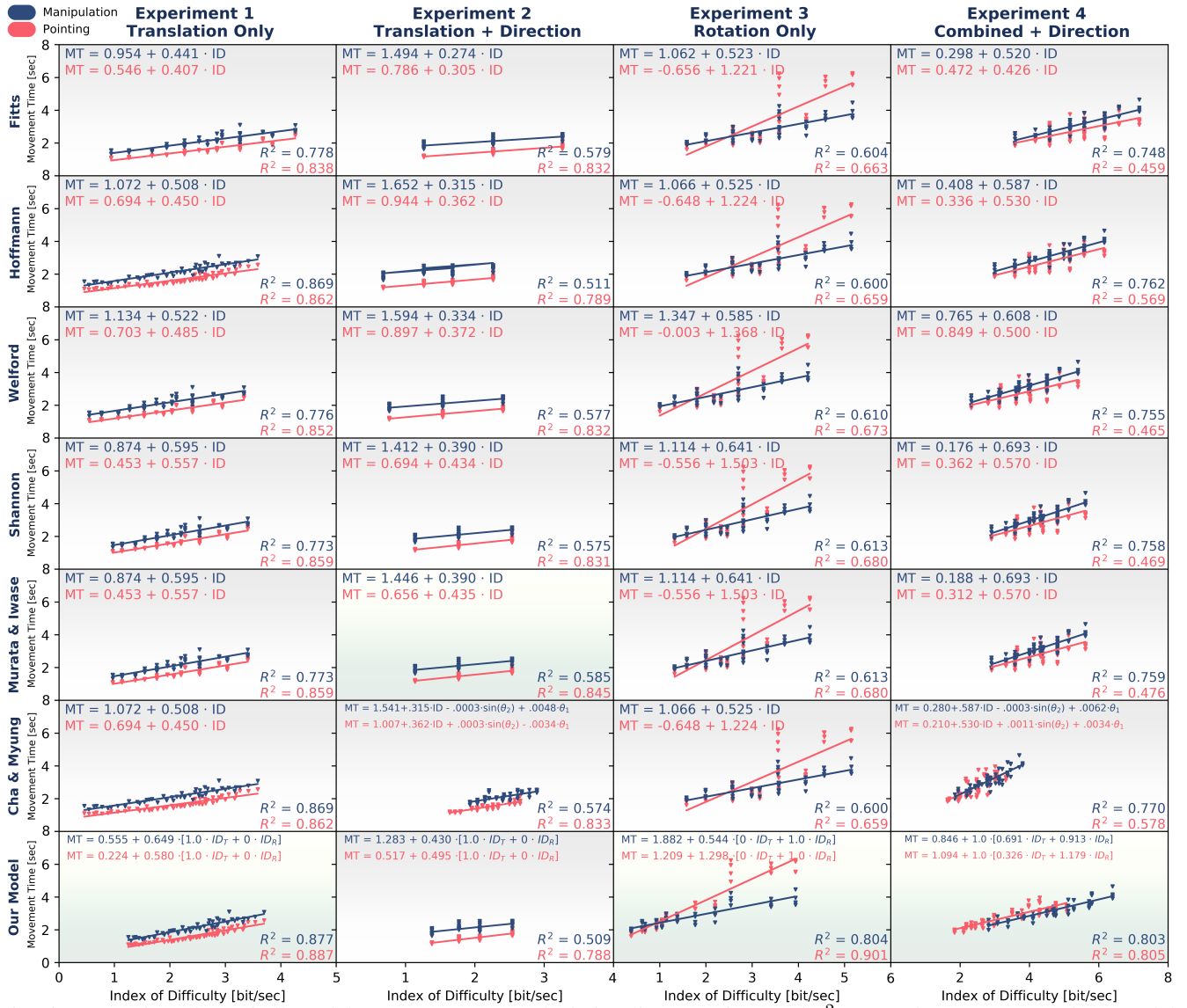
Fig. 6: Regression plots of all models across E1 to E4, depicting line equations and r^2. Green boxes represent best model.

direction ($0°$, $90°$), still investigating view and lateral directional influences. We also kept the object size fixed at $F = 4$cm. These design decisions were made to limit the number of tasks, allowing us to control our already exhaustive experiment.

1) Design: For the final experiment, a 2x2x2x2x2x2 within-subjects design was used, in which the independent variables were: two target sizes ($W = 4$ and 8cm), two target separations ($A = 12$ and 24cm), two direction angles ($\phi = 0°$ and $90°$), two inclination angles ($\theta = 15°$ and $30°$), two target rotations ($\alpha = 30°$ and $45°$) and two rotational tolerances ($\omega = 7.5°$ and $15°$). The dependent variable was MT. Representing 64 tasks and with 4 repetitions, we conducted 256 trials for both pointing and manipulation tasks, i.e. 512 trials per participant resulting in 10240 trials in total.

2) Pointing Results: An ART was applied since normality was violated ($p = .016$). The mean MT for pointing tasks was 2.71 ± 0.55s. A total of 120 error trials out of 4800, 2.5% of the data, were excluded from the analysis. Sphericity was met in all cases since all variations of the tested variables were two levels. There was no significant effect of target width W on MT, ($p > .05$). On the other hand, the target separation A, consistent with E1 and E2, significantly increased MT with higher separation values ($p < .001$). Neither directional arrangements ϕ nor inclination angles θ had a significant effect on MT, ($p = .141$) and ($p = .200$). As for rotation, α significantly affected MT ($p < .01$) and ω had the highest influence on MT ($p < .001$), consistent with E3.

3) Manipulation Results: The data was normally distributed ($p = .091$). The mean MT for manipulation tasks was 3.10 $\pm$ 0.59s. A total of 82 trials out of 4800, 1.7% of the data, were excluded from the analysis due to error. Sphericity was met in all cases. Contrary to the pointing task, target width W had a significant effect on MT ($p < .001$). Target separation A significantly increased MT with higher separation values ($p < .001$). Directional angles ϕ, as well as inclination angles θ, followed the same trend as with the pointing equivalent, not affecting MT much ($p = .099$) and ($p = .227$) respectively.

Rotational separation α did significantly affect the increase of MT as the degrees of separation rose ($p < .01$). Rotational tolerance ω had again the biggest influence on MT, being consistent with the pointing equivalent and E3 ($p < .001$).

4) Remarks: For E4, we added the IDs of translation and rotation of all existing approaches. The rotation was adjusted as described in E3 and in Table IV. Despite these models in principle being incompatible for combined movements, we added these two IDs to explore the potential of summing these two separate concepts of motion and to conduct a fair comparison. For example, the combined ID for Fitts' model would be $ID_t = log_2(\frac{2A}{W})$ for translation, plus $ID_r = log_2(\frac{2\alpha}{\omega})$ for rotation, with the other formulations following the same fashion. Overall, neither directional ϕ nor inclination θ angles had any significant effect on MT ($p > 0.05$) in pointing or manipulation. For combined movements, adding both IDs of translation and rotation, Fitts's, Shannon's, Welford's and even Murata & Iwase's directional formulation proved insufficient when fitted for pointing ($r^2 < 0.5$), but yielded slightly better results for manipulation ($r^2 < 0.77$). Hoffmann's had a significantly higher fitting towards pointing with an approximately 10% increase. Being consistent with E1, this suggests the necessity of adding the object size F in a model. Cha & Myung's had overall the highest fitting on the data but only marginally over the rest, primarily due to the incorporation of object size F as seen also in Hoffmann's.

VI. MODEL DERIVATION

In most prior work related with extending Fitts' law, a clear definition of distance is mostly missing [4]. For example, the distance from the object to the target (A) can either be defined from the object centre to the target centre (A_{cc}), the object centre to the target edge (A_{ce}) or even the object edge to the target edge (A_{ee}). Such a discrepancy highly influences the formula and can be formulated as:

$$A_n \begin{cases} A_1 = A_{cc}, \\ A_2 = A_{ec} + \frac{W}{2}, \\ A_3 = A_{ee} + \frac{W+F}{2}, \end{cases} \qquad d_n = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \tag{8}$$

From the above, a total of 3 different target separations are defined as A_n with $A_{cc} > A_{ec} > A_{ee}$. The Euclidean distance between two points in 3D is represented as d_n with $n = 3$. At each stage of the experiments, our results show that 3D translational movement follows Fitts' formulation closely. Target separation A and target width W are an integral part of the formulation. However, another variable with a determining effect in translational tasks was the object size F, as part of Hoffmann's model but not of Fitts'. Hence, the most appropriate definition of effective target separation would be A_3 from Eq. 8 which would include both W and F. By incorporating this definition into Fitts' model ID shown in Eq. 1, we have:

$$ID_t = log_2\left(\frac{2A}{F + W} + 1\right). \tag{9}$$

As for rotational movements, the graphs in Figure 5 show that while α follows an almost linear relationship with MT, and ω follows an evident non-linear relationship. While ω follows an almost inverse exponential decay, we believe a $\frac{1}{x^2}$ would be more appropriate, yet more data points would be needed to shed additional light on this discrepancy.

With the results at hand and the indicated discrepancy between translation and rotation as well as using the aforementioned distance definition, we can formulate an improved and more suitable 3D model as:

$$\text{Final Model:} \begin{cases} MT & = a + b[c \cdot ID_t + d \cdot ID_r], \\ ID_t & = log_2\left(\frac{2A}{F+W} + 1\right), \\ ID_r & = log_2\left(\frac{2\alpha}{\omega^2} + 1\right), \end{cases} \tag{10}$$

where a, b, c and d are constants determined through regression. We retain Fitts' original approach, and hence constants a and b are the y-intercept and slope respectively. Constants c and d represent the contribution of translation (ID_t) and rotation (ID_r) respectively, as these are two separate concepts of motion from our analyzed results. It is worthwhile to point out that the constant b is superfluous in this case since it multiplies constants c and d. However, the addition of the b constant allows the resemblance and familiarity of our model to that of Fitts' original formulation as seen in Eq. 1.

This model follows Stoelen & Akin's proposed formulation [11], and in particular the feasibility of adding the IDs of both translation and rotation into one combined index of difficulty. Yet in the formulation of [11], both ID_t and ID_r were controlled by a single constant b, as $MT = a + b \cdot [ID_t + ID_r]$. However, the assumption that translation and rotation have an equal contribution (i.e. 50% or 0.5) towards task difficulty (ID) and by extent MT is imprecise and limited.

In particular, our results show that rotation almost predominately affected MT and was primarily attributed to high rotational accuracy ($\omega = 2.5°$ in E3). For very small ω, we observed an almost inverse exponential decay with MT, instead of a linear one. Translation and rotation should hence be perceived as separate terms, and each should be quantified by its own constants. This mitigates the limitation of equal weighting as with [11], which is not suitable as shown in our study. In the following section, we analyze the results and compare the proposed model with the other aforementioned versions.

Model Fit (r^2)	Experiment 1 (Translation Only)		Experiment 2 (Translation w/ Directions)		Experiment 3 (Rotation Only)		Experiment 4 (Trans. w/ Rot. & Directions)	
	Pointing	Manipulation	Pointing	Manipulation	Pointing	Manipulation	Pointing	Manipulation
Fitts' [5]	0.838	0.778	0.832	0.579	0.663	0.604	0.459	0.748
Hoffmann's [8]	0.862	0.869	0.789	0.511	0.659	0.600	0.569	0.762
Welford's [17]	0.852	0.776	0.832	0.577	0.673	0.610	0.465	0.755
Shannon's [7]	0.859	0.773	0.831	0.575	0.680	0.613	0.469	0.758
Murata & Iwase [13]	0.859	0.773	**0.845**	**0.585**	0.680	0.613	0.476	0.759
Cha & Myung [12]	0.862	0.869	0.833	0.574	0.659	0.600	0.578	0.770
Our Model	**0.887**	**0.877**	0.788	0.509	**0.901**	**0.804**	**0.805**	**0.803**

TABLE IV: Overview of all methods with the respective r^2 fitting. For E3 and E4 where rotation is present, existing models are in theory incompatible. However, to retain consistency and to allow for a fair comparison, as motivated by [11], these were adjusted and extended by us to accommodate rotation as well. For example Fitts' $ID_t = log_2(\frac{2A}{W})$ would become $ID_r = log_2(\frac{2\alpha}{\omega})$ and both IDs are added for E4. Bold numbers represent the best method in the column group.

Research Implications and Findings	
R.1	Fitts' law can be extended towards 3D but does not adequately explain all spatial settings, and requires a clear definition of the distances associated in 3D.
R.2	Translation and rotation should be classified as separate concepts and quantities. Our analysis showed when rotation was present, it predominantly influenced MT over translation, by an almost $\frac{2}{3}$. Hence, a model has been built to account for these two separate concepts of motion by incorporating a constant for each, as these do not have an equal weight towards MT as observed by our experimental results.
R.3	Pointing and manipulation are perceived as inherently two very different types of tasks, but do follow and can be modelled by Fitts' law even in 3D space.
R.4	Object size does matter. During pointing but especially in manipulation this holds true. For the latter, the object size is an integral part of the metric. This is achieved by adding the object size F into the equation as suggested by Hoffmann [8] and confirmed in our model extension.
R.5	Rotation tolerance ω or simply the rotational accuracy, significantly affects MT. This is particularly the case for very low rotational tolerances ($2.5°$). Hence, strict rotational accuracy requirements during object placement (i.e. teleoperation) are likely to significantly increase operator completion timings.

Recommendations for Task Design in Pointing and Manipulation Tasks in VR and Teleoperation	
RC1	Avoid having "strict" spatial requirements towards accuracy. This holds especially true for rotation where very low rotational tolerances contribute to very high movement times, following an almost exponential relationship.
RC2	Avoid if possible in having pointing tasks that require movements either towards or away from the view direction of the user. Our results marginally showed that front-to-backwards movements take significantly longer than left-to-right ones and vice versa.

TABLE V: Summary of main results and key implications on object interaction in VEs and simulated teleoperation tasks.

VII. DISCUSSION

This study compared the most widely used extensions on Fitt's law through four designed experiments in a simulated teleoperation VR setting. This allowed us to evaluate the applicability of each model with various spatial complexities entailing translational and rotational movements. Each experiment included pointing and manipulation tasks, which are the most widely used types of interactions in collaborative VEs, VR simulators and robot teleoperation [1], [14], [16].

The study showed that in the most basic form of 3D object movement in a purely translational setting (E1), Fitts' law and its extensions are adequate for both target pointing and manipulation along a one-directional line only. However, when complexity increased by including directional and inclination angles, these models had reduced performance at predicting the results of our experiments, which dropped slightly in pointing tasks but significantly more in manipulation tasks (E2).

Though Fitts' law has been extended towards rotation, studies so far have limited their findings in 2D space [9], [10], [11]. Furthermore combining rotational and translational movements under one setting in 3D remains largely unexplored. While Kulik et al. [10] and Stoelen & Akin [11] studied combined movements, these were still limited to 2D space and only following movements across one line.

These current limitations in the state of the art showed their drawbacks in E3 and E4. All models in a purely rotational setting (E3) were insufficient when extended towards pointing and further aggravated during manipulation. However, the most important and critical experiment was E4, where we combined translational and rotational movements with varying spatial arrangements under one setting, rendering the last experiment a full 3D task. In E4, where we effectively investigated the most complex spatial settings in 3D, all compared models proved insufficient for both pointing and manipulation. These observations led us to the proposal of a new metric to overcome these limitations and the derived model outperformed other extensions in E1, E3 and most importantly in E4, as shown in Table IV and Figure 6.

Hence, our metric could be used to assess human movement in teleoperation [16] especially with the use of MR technologies [15]. We support this from our results, as our formulation can capture the complex spatial settings in 3D space associated with such scenarios. Moreover, as our formulation is based on Fitts' law and hence combines spatial and time-based metrics under a single model, further studies on HCI/HRI could benefit in terms of standardization [2], [3].

A. *Interpreting Results on Combined Movements in 3D*

For combined movements, even by adding the IDs of translation and rotation (Stoelen and Akin [11], and Kulik et al. [10] in 2D tasks), the model is insufficient. The latter study observed only a linear fitting of $r^2 = .780$ when combining both translation and rotation in simple 2D settings following movements across one line. Yet direct comparisons between

different studies are difficult, particularly due to correlation coefficients (r^2) being highly influenced by the involved data points. Consequently, different experimental settings and manifestations do not allow for inter-study comparability [2], [4]. Hence, we compare these models on our data directly across four experiments, as summarized in Table IV and Figure 6.

By comparing the most widely used model extensions to date, our proposed model demonstrated better fitting results than the existing extensions both in E1 and E3, and most importantly when both translational and rotational movements were combined (E4). In experiment E4, our model was better fitting for both pointing ($r^2 = .805$) and manipulation ($r^2 = .803$) than any other formulation based on Fitts' law. This result can be attributed to three major factors.

Firstly, our results showed that equally weighing both IDs of rotation and translation and by extent assuming that both have a linear effect on MT, is not supported by our analysis. Instead, it is crucial to separate these two terms, which can be accomplished by introducing different constants for each as explained in Section VI. Secondly, it shall be noted that the rotational tolerance ω has a non-linear relationship with MT in our study, in contrast to a linear relationship as in [11]. This difference implies that higher degrees of difficulty require matching the rotation of an object in all 3D axes, instead of one as in their work, limited to 2D space [11]. Thirdly, we can also conclude that from the 2D formulations, Hoffmann's formulation [8] indeed works well and indicates the necessity of adding the size of the object F, hence why it is included in our model. We can thus infer that both W and F inversely affect MT and should therefore be included in a model.

Our experiments found no evidence to support the significance of either directional or inclination angles in a consistent manner. More specifically we observed that directional angles presented only a significant influence towards MT limited in E2 and only in pointing ($p < .05$). Furthermore, while inclination angles affected MT both for pointing and manipulation in E2 ($p < .05$), they presented an insignificant effect in E4 with combined movements. In E4, the translational and rotational variables influenced MT significantly more than inclinations or directions. Nevertheless, despite its negligible effect on MT, our study shows marginally worse performance when objects were presented along the view axis i.e. front-to-back ($90°$, $270°$) rather than the lateral i.e. left-to-right ($0°$, $180°$). This is in line with [1].

Finally, manipulation tasks took a longer time to complete in all cases except in purely rotational tasks (E3), where pointing proved significantly more difficult, especially with the lowest target tolerance at $\omega = 2.5°$. Overall, this time difference can be partially explained by the duration spent on grasping the object. An important finding was that manipulation can be modelled under Fitt's model if the important variable of object size F is taken into account. The object size is mostly overlooked in the state of the art but is mitigated in Hoffmann's work [8], as seen in Eq. 2.

In Table V, we summarize the main findings from our work and the implications on task design involving pointing tasks, 3D user interfaces as well as the manipulation of objects in 3D space, as seen in VR and robotic teleoperation.

B. Limitations and Future Work

Some limitations to overcome in a future study is to consider the multitude of different input devices and different grasping types one can use, which may significantly influence the formulation of human models [2]. Furthermore, human perception is a complex phenomenon and subject to each individual's exposure to technologies as well as personality-related factors, yet accounting for these is particularly challenging [2].

In this work, we investigated seven distinctive spatial variables to explain and model combined translational and rotational variations in full 3D space, as shown in Table II. Consequently, due to the significant amount of variations introduced, we limited the investigation of each variable to a maximum of four levels. Future studies are advised to investigate an even wider range of these variables with more levels, to further increase our understanding of their contribution towards task difficulty. For example, changing the cube shape in our study, with a more complex 3D shape such as a toy car, will allow for a wider range of rotational variations to be investigated.

As for all human performance models deriving from Fitts' law, the main limitations are stationary pointing or manipulation tasks. All models studied in this paper do not take into account upper-body movements, such as movements from torsos and/or shoulders. Furthermore, biomechanical variations in participants, such as placing objects near the dominant hand or the size of the controlled end-effector are all determining factors for a model.

VIII. CONCLUSION

In this work, we investigated a new human performance metric in full 3D for both pointing and manipulation with combined translational and rotational movements. We conducted four experiments, each adding progressively higher spatial complexity. In the most basic form of 3D translational pointing and manipulation (E1), we observed that existing approaches can be used to adequately model human performance, but were insufficient when spatial arrangements were introduced (E2), such as inclinations and directions. However, the majority of these approaches did not model well the rotational movements as shown in E3 and were insufficient when combining translation and rotation with spatial arrangements as studied in E4.

To the best of our knowledge, this study extensively compared the most widely used performance metrics based on Fitts' law. We proposed a new performance model that can model human performance in full 3D space in a simulated teleoperation VR setting with two types of interactions. Our metric modelled human performance better than existing formulations, especially in increasing spatial complexities such as those seen in full 3D space.

However, numerous aspects should further be investigated and improved, such as a more comprehensive range of spatial variations (e.g. additional levels of rotations, spatial directions etc.), more complex shapes (non-primitive 3D objects), different grasping types, hand end-effectors as well as input devices. Nonetheless, this work was a first attempt towards a higher dimensional formulation to evaluate full-3D movements, which can be used in accessing human performance in a VR-based teleoperation setting and with the potential to be applied in robot shared control.

ACKNOWLEDGMENTS

REFERENCES

[1] M. D. Barrera Machuca and W. Stuerzlinger, "The effect of stereo display deficiencies on virtual hand pointing," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3290605.3300437

[2] E. Triantafyllidis and Z. Li, "The challenges in modeling human performance in 3d space with fitts' law," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3411763.3443442

[3] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, ser. HRI '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 33–40. [Online]. Available: https://doi.org/10.1145/1121241.1121249

[4] H. Drewes, "Only one fitts' law formula please!" in *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 2813–2822. [Online]. Available: https://doi.org/10.1145/1753846.1753867

[5] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement." *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.

[6] P. M. Fitts and J. R. Peterson, "Information capacity of discrete motor responses." *Journal of experimental psychology*, vol. 67, no. 2, p. 103, 1964.

[7] I. S. MacKenzie, "Fitts' law as a research and design tool in human-computer interaction," *Hum.-Comput. Interact.*, vol. 7, no. 1, p. 91–139, Mar. 1992. [Online]. Available: https://doi.org/10.1207/s15327051hci0701_3

[8] E. R. Hoffmann, "Effective target tolerance in an inverted fitts task," *Ergonomics*, vol. 38, no. 4, pp. 828–836, 1995. [Online]. Available: https://doi.org/10.1080/00140139508925153

[9] G. V. Kondraske, "An angular motion fitt's law for human performance modeling and prediction," in *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, 1994, pp. 307–308 vol.1.

[10] A. Kulik, A. Kunert, and B. Froehlich, "On motor performance in virtual 3d object manipulation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 2041–2050, 2020.

[11] M. F. Stoelen and D. L. Akin, "Assessment of fitts' law for quantifying combined rotational and translational movements," *Human Factors*, vol. 52, no. 1, pp. 63–77, 2010, pMID: 20653226. [Online]. Available: https://doi.org/10.1177/0018720810366560

[12] Y. Cha and R. Myung, "Extended fitts' law for 3d pointing tasks using 3d target arrangements," *International Journal of Industrial Ergonomics*, vol. 43, no. 4, pp. 350 – 355, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169814113000723

[13] A. Murata and H. Iwase, "Extending fitts' law to a three-dimensional pointing task," *Human movement science*, vol. 20, no. 6, pp. 791–805, 2001.

[14] E. Triantafyllidis, C. Mcgreavy, J. Gu, and Z. Li, "Study of multimodal interfaces and the improvements on teleoperation," *IEEE Access*, vol. 8, pp. 78 213–78 227, 2020.

[15] S. Fani, S. Ciotti, M. G. Catalano, G. Grioli, A. Tognetti, G. Valenza, A. Ajoudani, and M. Bianchi, "Simplifying telerobotics: Wearability and teleimpedance improves human-robot interactions in teleoperation," *IEEE Robotics Automation Magazine*, vol. 25, no. 1, pp. 77–88, 2018.

[16] J. M. O'Hara, "Telerobotic control of a dextrous manipulator using master and six-dof hand controllers for space assembly and servicing tasks," *Proceedings of the Human Factors Society Annual Meeting*, vol. 31, no. 7, pp. 791–795, 1987. [Online]. Available: https://doi.org/10.1177/154193128703100723

[17] A. T. Welford, *Fundamentals of skill*, ser. Methuen's manuals of modern psychology. London: Methuen, 1968.

[18] Y. Wang and C. L. MacKenzie, "Object manipulation in virtual environments: Relative size matters," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '99. New York, NY, USA: Association for Computing Machinery, 1999, p. 48–55. [Online]. Available: https://doi.org/10.1145/302979.302989

[19] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, 2019. [Online]. Available: https://science.sciencemag.org/content/364/6446/eaat8414

[20] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, "The aligned rank transform for nonparametric factorial analyses using only anova procedures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 143–146. [Online]. Available: http://doi.acm.org/10.1145/1978942.1978963