Edinburgh Research Explorer

The Challenges in Modeling Human Performance in 3D Space with Fitts' Law

The Challenges in Modeling Human Performance in 3D Space with Fitts' Law

Eleftherios Triantafyllidis
School of Informatics, The University of Edinburgh
Edinburgh, United Kingdom
eleftherios.triantafyllidis@ed.ac.uk

Zhibin Li
School of Informatics, The University of Edinburgh
Edinburgh, United Kingdom
zhibin.li@ed.ac.uk

ABSTRACT

With the rapid growth in virtual reality technologies, object interaction is becoming increasingly more immersive, elucidating human perception and leading to promising directions towards evaluating human performance under different settings. This spike in technological growth exponentially increased the need for a human performance metric in 3D space. Fitts' law is perhaps the most widely used human prediction model in HCI history attempting to capture human movement in lower dimensions. Despite the collective effort towards deriving an advanced extension of a 3D human performance model based on Fitts' law, a standardized metric is still missing. Moreover, most of the extensions to date assume or limit their findings to certain settings, effectively disregarding important variables that are fundamental to 3D object interaction. In this review, we investigate and analyze the most prominent extensions of Fitts' law and compare their characteristics pinpointing to potentially important aspects for deriving a higher-dimensional performance model. Lastly, we mention the complexities, frontiers as well as potential challenges that may lay ahead.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Interaction techniques**; *HCI theory, concepts and models*; *Gestural input*; *HCI design and evaluation methods*; Graphical user interfaces; Virtual reality; Mixed / augmented reality; Pointing.

KEYWORDS

Fitts' Law, 3D Performance Model, Human Performance, 3D Pointing, 3D Manipulation, Motor Performance Review, Object Interaction, Virtual Reality, Models and Metrics, Evaluation Metrics, Gestural Input, Interaction Techniques

1 INTRODUCTION

Fitts' law is the most widely used human performance model in HCI history [4, 17, 24, 31, 40, 52, 76]. With the growth of networking and mixed reality technologies, the necessity of assessing human performance with a higher dimensional model has increased significantly [17, 82]. Fitts' proposed his original model, known as simply Fitts' model in 1954 [19]. Its applicability has been well demonstrated in 2D tasks [19, 32, 52], motivating multiple other researchers with their own 2D extensions, most notably that of Hoffmann's, Welford's and Shannon's [52, 87, 90]. More recently, the law itself has also been tested in 3D space [44, 77] and providing a robust ground for various 3D extensions [5, 14, 59], with lesser but still impressive predictive powers.

It would thus come naturally that somewhere out there, there would exist an extension of Fitts' law to model the entirety of the 3D domain. However, to the best of our knowledge, this is not the case. The most popular methods to date have provided an invaluable insight as to which variables are integral when spatially describing 3D space [5, 14, 44, 52, 59, 77, 87, 90]. Consequently, these methods were tested under very specific settings and limited to certain spatial arrangements, thus progress towards a true 3D performance metric is still scattered not allowing for inter-study comparisons. More specifically, a few models accounted for varying gains of spatial arrangements including directions and inclinations [5, 6, 14, 59], of which only the latter two formulated an extension. While Fitts' model was originally intended for translational tasks only, other studies demonstrated that it can adequately model 2D rotational tasks as well [15, 41, 42], though not exhaustively tested in the 3D domain. Furthermore, combined translational and rotational movements are severely limited and only accounted by two studies [44, 77], of which movements were limited only across one line, effectively disregarding spatial arrangements. Moreover, depth-related distances in 3D displays were only accounted for and added to a model extension in one study [5]. Last but not least, with the exception of Hoffmann's [32] all aforementioned studies only considered the effective target size, effectively ignoring the probe size i.e. the object size used to point to the target location. All of these studies reported significant influences to their models, yet with the absence of a standardized metric, inter-study validations remain particularly challenging.

We can thus infer that there are numerous spatial variables and multiple complexities arising when trying to extend Fitts original model to 3D space. With this review, we hope to investigate and pinpoint towards important factors one has to consider if attempting to propose a model in 3D space, taking into account varying gains of translational and rotational complexities including spatial arrangements.

2 BACKGROUND AND RELATED WORK

In this section, we investigate and analyze the most prominent and widely used extensions based on Fitts law. Moreover, we compare all formulations and their applicability in increasing spatial complexities as to pinpoint towards which directions researchers should direct focus to if attempting to derive a full 3D performance model. In Table 1 we summarize all models and extensions as an overall visual overview of their applicability towards higher dimensions.

2.1 Original Law

Fitts' original formulation [19, 20] predicts the *movement time* (MT) based on an index of difficulty (ID) and is formulated as follows:

$$MT = a + b \cdot ID,$$
$$ID = log_2\left(\frac{2A}{W}\right) \tag{1}$$

It can be thought of as the time to reach/point or click to a target location, given the target's distance (A) from the origin of the cursor/hand or object, as a ratio of the target's width (W). The logarithmic term *ID*, represents the *index of difficulty* measured in bits per second [bit/s] while *MT* is measured in seconds. Constants *a* and *b* represent the intercept and slope respectively and are derived via regression. In the following sections, we will present the most widely used extensions of Fitts' model, also indicating the ID for each equation to make it clear; as with most work, it is not clearly stated which part represents the actual ID.

2.2 Extensions in 2D Space

Including Fitts' formulation, there are numerous variants based on his extensions, including:

$$MT = a + b \cdot \underbrace{log_2\left(\frac{2A}{W}\right)}_{ID} \tag{2}$$

$$MT = a + b \cdot \underbrace{log_2\left(\frac{2A}{W+F}\right)}_{ID} \tag{3}$$

$$MT = a + b \cdot \underbrace{log_2\left(\frac{A}{W}+0.5\right)}_{ID} \tag{4}$$

$$MT = a + b \cdot \underbrace{log_2\left(\frac{A}{W}+1\right)}_{ID} \tag{5}$$

Equation 2 represents Fitts' original formulation. Equation 3 is an extension of Fitts' by Hoffmann [32]. In his formulation, Hoffmann kept Fitts' law mostly intact, with the exception of adding the variable *F*, representing the index finger pad size of each participant. This stemmed from a series of experiments he conducted mainly composed of discrete tapping tasks using the participants' pad finger size of their index as pointing probes. His formulation opened new and interesting paths. Indirectly, it motivated future work in including the object size when concerned with manipulating objects in virtual environments [87]. Equation 4 presents Welford's

extension [90], removing the multiplication by 2 for the target separation i.e. distance (A) but adding +0.5 in his formulation. Similarly, in Equation 5 Scott MacKenzie introduced the Shannon's formulation [52], similarly to Welford's but instead of 0.5 adding a plus +1 term. The difference between the latter two lays in the added terms. Welford, contrary to MacKenzie, argued that the reason for adding the +0.5 term was to account for the distance from the centre of the target to its edge. Based on the Shannon Formulation [52], another model extension, named FFitts Law [7], was suggested and formulated as:

$$MT = a + b \cdot \underbrace{log_2\left(\frac{A}{\sqrt{2\pi e\left(\sigma^2 - \sigma_\alpha^2\right)}}+1\right)}_{ID} \tag{6}$$

replacing the denominator, effective target width (W), with a double Gaussian distribution in which σ represents the standard deviation of the touch points and σ_α the precision of the input finger. This approach showed a promising accuracy in both 1D and 2D target acquisitions tasks.

Nonetheless, the most widely used 2D extension of Fitts law still remains that of MacKenzie's i.e. the Shannon formulation and has been demonstrated to do very well for tasks entailing purely translational [14, 27, 33, 34, 53, 72, 77] or rotational settings [58, 77].

2.3 Extensions in 3D Space

Fitts' formulation has also been applied in the 3D domain but has shown not to represent 3D movements accurately [6, 37, 75, 77, 95]. Consequently, certain extensions were needed to account for this limitation [5, 14, 29, 43, 59, 64]. The first being directions. Murata and Iwase [59], were the first to introduce directional angles in their study, in their case phrased as azimuth angles under the spherical coordinate system. A total of eight different levels of directional angles were investigated ranging from 0° to 315° with a 45° increment. Their findings showed that these angles had a sinusoidal relationship with movement time. More specifically, they found that upper (90°) and lower movements (270°), were significantly more difficult and by extent increasing MT than left (180°) or right movements (0°). Cha and Myung [14], inspired by Murata and Iwase's model [59], proposed an additional term, inclination. In their experiment, they introduced both varying gains of directional as well as inclination angles for pointing tasks. They confirmed Murata and Iwase's work in that directional angles do indeed appear to have a sinusoidal relationship with MT and moreover, inclination angles appeared to have an almost linear relationship with MT. Neither of those models however investigated their findings within an experimental setting of 3D displays, let alone VRHMDs. This was later accounted for by Machuca and Stuerzlinger [5]. The latter investigated the stereo deficiencies in virtual hand pointing with the use of 3D displays. Foremost, they confirmed that left to right movements were significantly easier than movements away from or towards the user. However, it is important to mention that both Murata & Iwase [59], as well as Cha & Myung [14], studied this discrepancy with the directional azimuth angles being perpendicular to the view direction of the participant, i.e. a frontal circle in front of them. Whereas in Machuca and Stuerzlinger's,

the directional angles were placed around the participant with $90°$ and $270°$ representing the front and backward whereas $0°$ and $180°$ degrees represented left and right movements respectively. The most important however finding was that depth changes i.e. the distance of the user's eyes to the screen, linearly affected MT, with higher depth values presenting higher difficulties and by extent higher timings. The aforementioned models are shown below.

$$MT = a + b \cdot \underbrace{\left(log_2 \left(\frac{D}{W} + 1 \right) + c \cdot \sin \theta \right)}_{ID} \qquad (7)$$

$$MT = a + b \cdot \theta_1 + c \cdot \sin \theta_2 + d \cdot \underbrace{log_2 \left(\frac{2D}{W + F} \right)}_{ID} \qquad (8)$$

$$MT = a + b \cdot \underbrace{log_2 \left(\frac{A}{W} + 1 \right)}_{ID} + c \cdot CTD \qquad (9)$$

Murata and Iwase's [59] directional model is shown in Equation 7. The variable $\sin \theta$ represents the sinusoidal directional / azimuth angle, controlled by a constant c, determined through regression. Note that in Murata and Iwase's model, contrary to other extensions, the ID not only encompasses the logarithmic but also the sinusoidal term with the azimuth angle added to it. Cha and Myung's model is shown in Equation 8. Contrary to Murata and Iwase's model which is based on Shannon's, Cha and Myung's [14] is based on Hoffmann's taking into account the finger pad size of the participants acting as the pointing probe (F). In addition to the directional angle ($\sin \theta_2$), they introduced θ_1 which represents inclinations sharing a linear relationship with MT. Constants a,b,c and d are again determined through regression. Finally, Machuca and Stuerzlinger's model [5] is shown in Equation 9, which is based on Shannon's [52] shown in Equation 5 with the addition of CTD representing the Change in Target Depth (measured in centimeters), controlled by a constant c determined through regression. None of the aforementioned models in this section, however, included combined translational or rotational variations.

2.4 Translation or Rotation?

To this point, all the formulas reported, either extended the original Fitts' law from 2D to 3D space, but were solely limited to translation. However, during object interaction, be it pointing or manipulation, rotation is a fundamental part. When performing a task that requires some kind of spatial accuracy, we humans usually attempt to match the rotation of the object so it satisfies certain spatial criteria [82].

Stoelen and Akin [77], were the first to combine both translational and rotational movements in one experiment. Motivated by MacKenzie's Shannon formulation [52] shown in Equation 5, Stoelen and Akin proposed that simply adding the indices of translation and rotation would yield adequate results in modeling combined movements. To adjust the formula of translational movements of Shannon's, Stoelen and Akin replaced the otherwise target distance (A) as the numerator, with the respective rotational distance (α) and the denominator (W) indicating the target width, with the rotational tolerance (ω). Their formulation is shown in Equation 10:

$$MT_{combined} = a + b \cdot (ID_{translation} + ID_{rotation})$$
$$ID_{translation} = log_2 \left(\frac{A}{W} + 1 \right) \qquad (10)$$
$$ID_{rotation} = log_2 \left(\frac{\alpha}{\omega} + 1 \right)$$

As such, the total combined movement time ($MT_{combined}$) it takes to point to a target is dependent upon the index of difficulty for translation ($ID_{translation}$) and rotation ($ID_{rotation}$), sharing the same "weight" and linearly correlated to MT. Stoelen and Akin, however, limited their findings without any spatial arrangements since pointing to the target was performed only across one line and furthermore performed in 2D space without the use of 3D displays [77]. The latter was accounted for by Kulik et al. [44], whereby 3D displays were used to model combined transitional and rotational movements. They found a surprisingly adequate linear fit of $R^2 = 0.78$ when combining both translation and rotation, defined in their case as "3D Docking" [44]. They confirmed the findings of Stoelen and Akin yet movements were again limited along one line only. The question however remains, does a rotation in 3D space really share a linear relationship with MT as with translation? With the exception of the two studies mentioned above [44, 77], this is not sufficiently investigated and perhaps rotation could share a polynomial or exponential relationship with MT. It would thus be invaluable to further confirm their findings.

2.5 Pointing or Object Manipulation?

In the larger context of object interaction, there are in general two major categories, pointing and manipulation. Until now, we investigated purely pointing tasks, that is without the presence of any physical interactions such as gravity or contact points. One could argue that covering pointing tasks under a unified 3D performance model would suffice. However, use cases encompassing teleoperation and in general simulation training of operators, heavily depend upon as close to real physics as possible [45, 82]. Physical interactions can be perhaps ignored in 3D user interfaces up to a point, yet with the vast availability of mixed reality technologies on the market and increasingly more powerful hardware, the need to model physics properties has significantly increased. Object manipulation even in multi-user environments becomes more and more popular and interaction users initiate with the environment should inherently include physical properties otherwise these may be perceived as breaking immersion [61].

Sadly, the application of Fitts' law towards manipulation is severely limited [23], particularly due to being intended for pointing tasks in the first place. Yet, if we break down the phases of pointing and manipulation, we can see that these do not differ that much from another. Nieuwenhuizen [62] studied and proposed the phases observed in 3D goal-directed movements. He proposed five phases that are generally seen when interacting with objects, latency, initiation, ballistic, correction and verification phase. During the first two phases, the velocity of hand movements is minimal, while during the ballistic phase it is at its maximum. During the correction and verification phase, velocity drops as users correct any object errors such as increasing accuracy of placements. This should not differ for either pointing or manipulation. For simplicity purposes,

both pointing and manipulation can, in essence, be broken down into three at minimum parts: (a) acquisition or grasping phase, (b) transportation phase and (c) correction phase. The first phase (a) would merely differ in its name depending on either pointing or manipulation while (b) and (c) would, in essence, be similar.

The closest work that investigated the applicability of Fitts' model for manipulation tasks is that of Yanqing and L. MacKenzie [87]. While not introducing a new model, they concluded that object size, similar to Hoffmann's model [32] shown in Equation 3 greatly affects MT, with bigger dimensions corresponding to improved performance i.e. lesser MT and also linearly correlated with time. In Table 1 below, we summarize all model equations investigated thus far, including their characteristics and applicability under different spatial settings.

3 OUTLOOK, CHALLENGES AND FRONTIERS

To this point, we investigated and analyzed the most prominent and widely used extension of Fitts' law. However, there are still numerous challenges to address if one wants to propose a unified 3D model covering the entirety of the 3D domain. Hence in this section, we will briefly go over the challenges, frontiers and general outlook of what researchers should expect when deriving a standardized human performance metric in full 3D space. Finally, in Table 2 we summarize potentially important research directions, aims and questions as the result of this review including sources and readings for other researchers to pursue.

3.1 Influences of Depth Perception

The main limitation of applying Fitts' model in 3D pointing, especially with the use of Mixed Reality (MR) technologies is mostly attributed to impaired depth perception [6, 35, 51]. Numerous studies support that the estimation of distances for virtual targets differs to that of physical targets, in that humans appear to overestimate their ability to perceive depth in virtual environments (VEs) and by extent the target depth to reach [47, 68, 78, 91]. One study even estimated that this discrepancy differed with the real target to an almost 74% of their true distance [68]. Other studies argued that to overcome depth limitations, one could increase the display resolution of 3D displays to provide higher representations of depth, potentially implying that this may mitigate to some extent distance overestimation [25, 39]. As mentioned, Machuca and Stuerzlinger investigated and proposed a simple model accounting for depth perception in VEs. Their main finding was that movements along the depth axis i.e. away or towards the user appears to be more difficult than left to right movements, which appears to be supported by earlier work [14, 59, 74]. Including as such the effect of depth distances and adding these to a model is vital and with Machuca and Stuerzlinger's work, it does appear to have a significant effect [5]. Whether it shares a linear relationship as with the latter, needs more verification and testing since studies investigating the effects of depth perception on Fitts' law is still limited.

3.2 Evaluation Approaches

A popular approach in evaluating Fitts' model and its extensions, is to use the coefficient of determination (R^2) to assess the correlation between the ID and MT drawn on the x and y-axis respectively.

The closer this value is to 1, the "better" the fit between these two variables and closer to 0 explaining "less" in return. Generally, an extension is deemed to be "superior" when tested against other models when there is a better correlation. However, there are numerous disadvantages to merely using the (R^2), which is supported by current literature [17, 26].

The major disadvantage of merely reporting the coefficient of determination is that it is highly dependent upon the number of data points recorded. More specifically, the more data pairs there are in the evaluation, the correlation will usually be lower. This can be exploited by having a small number of IDs and a lot of repetitions of the same tasks to achieve relatively "easily" a good resolution.

Heiko Drewes [17] illustrated this limitation by presenting the difference of the (R^2) results in a single click-the-target experiment versus the same experiment but repeated multiple times and averaged over. Motivated by his observation, we took the same notion and applied it in our case as well with more levels of repetitions. We conducted a simple pointing task with four target sizes (W = 5, 7.5, 10, 12.5 [cm]) and four target separations (A = 12, 24, 36, 48 [cm]). Amounting to 16 distinctive tasks. By applying Fitts' law shown in Equation 1, a total of 16 IDs were calculated, ranging from 0.941 to 4.26 bits. A total of 20 repetitions were made by a single participant. Figure 1 visually illustrates the reported R^2 values in the 3D pointing task of a single repetition versus 5, 10 and 20 repetitions. Notice that the more repetitions we have, the higher the R^2 value becomes.

Unfortunately, most literature to date use this evaluation approach exclusively [6, 14, 44, 59, 77]. While it does indeed provide an overview of how well the model does, researchers would also be advised to report the values of constants and their respective standard error as shown in Figure 1. For example, merely stating that a model has a correlation of $R^2 = 0.906$ is not very helpful. Instead, it would be much more useful to not only provide the latter, but also include a statement such as "Fitts' original formulation showed a correlation of $R^2 = 0.906$ with MT = 0.66 (0.07) [0.52, 0.79] + 0.27 (0.02) [0.21, 0.31] · ID". The terms represent the a and b constants respectively, reporting the standard error in addition to the lower and upper bound confidence intervals (ideally CI=95%) that can be reported through regression analysis.

Other methods that are also used to predict the differences between the MT and ID for a particular model and its fitting, is the Root-Mean-Square Deviation (RMSD) as well as the Mean Absolute Deviation (MAD) [38]. Solely relying on the R^2, or for any statistical model in that case, as we saw has numerous limitations, particularly limiting inter-study comparability. Perhaps a better way for authors to support their argumentation is a mixture of multiple statistical models to show their model fitting.

3.3 Human Factors and Experimental Design

By definition, Fitts' prediction model, models the performance of humans. For example, human performance is dependent upon ones own personality-related factors including but not limited to age, visual health, previous exposure to certain technologies, absorption as well as cognitive ability [30, 56, 57, 71, 94]. As such, factors such as tiredness, cognitive ability, concentration, being under the

Human Performance Models	Model Formulation		Model Characteristics				
	MT	ID_t / ID_r*	Based On	Space	Dir.*	Inc.**	Depth
Fitts' [19]	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{2A}{W}\right)$	N/A	2D	No	No	No
Shannon's [52]	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{A}{W} + 1\right)$	[19]	2D	No	No	No
Hoffmann's [32]	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{2A}{W+F}\right)$	[19]	2D	No	No	No
Welford's [90]	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{A}{W} + 0.5\right)$	[19]	2D	No	No	No
Murata and Iwase's [59]	$MT = a + b \cdot ID$	$ID_t = log_2\left(\frac{A}{W} + 1\right) + c \cdot \sin\theta$	[52]	3D	Yes	No	No
Cha and Myung's [14]	$MT = a + b \cdot \theta_1 + c \cdot \sin\theta_2 + d \cdot ID$	$ID_t = log_2\left(\frac{2A}{W+F}\right)$	[32, 59]	3D	Yes	Yes	No
Stoelen and Akin's [77]	$MT = a + b \cdot [ID_t + ID_r]$	$ID_{t/r} = log_2\left(\frac{A_t/\alpha_r}{W_t/\omega_r} + 1\right)$	[52]	3D	No	No	No
Machuca and Stuerzlinger's [5]	$MT = a + b \cdot ID + c \cdot CTD$	$ID_t = log_2\left(\frac{A}{W} + 1\right)$	[52]	3D	No***	No	Yes

Table 1: Summary of the most widely used 2D and 3D extensions of Fitts' law. Table illustrates the equations as defined by the respective authors. Model characteristics represent the model settings and whether these are covering important spatial characteristics to model full 3D performance. $ID_{t/r}$: Index of difficulty of translation (*t*) or rotation (*r*). Dir.*: Directions. Inc.: Inclines. No***: Effects were investigated but no formulation or model extension was performed. [N/A]: Not applicable.**
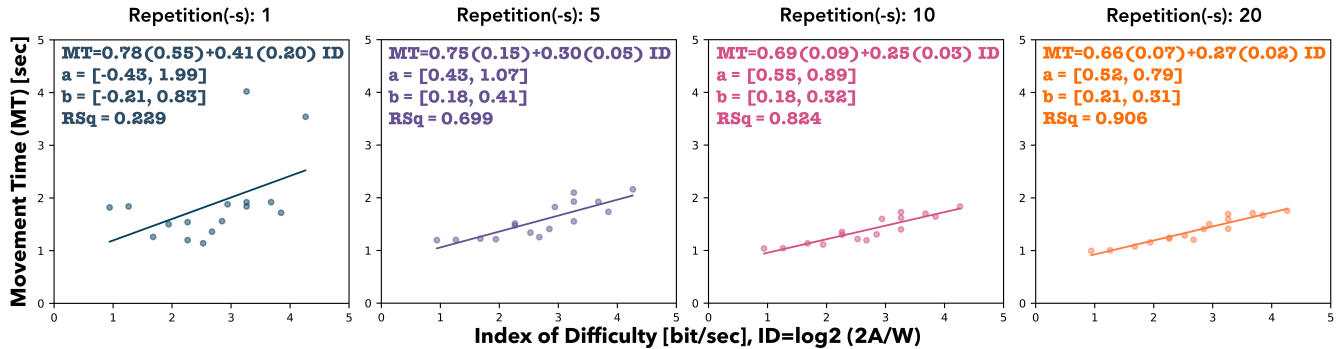

Figure 1: An example of how 1,5,10 and 20 repetitions of a simple point to the target task affect linear regression and more specifically the influence of (R^2). Notice the significant increase of the (R^2) for repetition 1, 5, 10 and 20 corresponding to $R^2 = 0.229$, $R^2 = 0.699$, $R^2 = 0.824$ and $R^2 = 0.906$ respectively. This figure serves as an example of a "more" helpful overview of the model fitting results between the ID and MT should be. More specifically we report the full line equation $MT = a(SE) + b(SE) \cdot ID$, with ($SE$) representing the standard error and reporting the lower and upper bound confidence intervals of the constants (CI=95%). We advise researchers to follow a similar approach instead of merely focusing on and reporting the R^2 values.

influence of stimulants etc. may have a determining effect on performance. While perhaps modelling and accounting for these factors in a formulation may be significantly challenging, studies should limit or retain consistency when recruiting participants. For example, all participants in a study should ideally be evaluated if they are tired or are under the influence of certain stimulants e.g. coffee, and those who do not meet the criteria should be excluded from the analysis or initial recruitment. Sadly, most studies so far miss a clear definition of the state of the participants [14, 44, 59] as also suggested by Heiko Drewes [17]. Consequently, more focus should be given when reporting participant recruitment.

Another factor to take into account is the specific experimental design that each author considered to this date when deriving a model extension. Merely deriving a human performance model under one experimental setting is insufficient as the authors are likely in the risk of observing and tuning a model only to fit that particular study manifestation. Ultimately, if a model is only applicable for very specific application settings, its relevance can become questionable and likely at risk of losing overall generalizability. These limitations can be mitigated to some extent if one would collectively include all models as seen in Table 1, in addition to any subsequent ones, and compare them under one particular 3-D experimental design setting to support their argumentation.

3.4 Potential Benefits of Multimodality

Multimodal interfaces can mitigate the high complexities that inherently surround interaction in VEs. It is generally argued that human interaction with the surrounding environment is inherently multi-modal [10, 67, 84]. More specifically, manipulation by itself is multi-modal, requiring more than one modality to be active to effectively control, approach and grasp an object [8].

Research Considerations and Questions to Explore		Sources & Readings
Q1	Explore the effect of object size as suggested by Hoffmann in the context of both pointing and manipulation with more levels and a more exhaustive setting.	[33, 87]
Q2	In Hoffmann's formulation, can we assume that the object size (F) can represent the dimensions of a given 3D object as it does in representing the pad size index finger? Are we likely in need of different definitions?	[33, 87]
Q3	How do different input technologies affect a given model? Can it be accounted for by implementing an additional term in the formulation or controlling existing terms with additional constants?	[12, 55, 72]
Q4	Spatial arrangements, directions & inclinations, appear to matter according to Murata & Iwase, Cha & Myung and Machuca and Stuerzlinger, but can their findings be confirmed and furthermore extended in combined translational and rotational movements? Potentially of significant importance due to limited focus.	[5, 14, 59, 85]
Q5	Is the simplicity of adding two indices of difficulty of translation and rotation feasible? Can rotation indeed be modelled by merely replacing the target distance (A) with the rotational offset (α) and the target width (W) with the rotational tolerance (ω)? For translational tasks, A and W have been observed by numerous studies to share a linear relationship with MT, is that the case for the less explored rotational counterpart? Does rotation perhaps share a polynomial or even exponential relationship with MT? The aforementioned should be answered and part of these are only explored and limited to two studies by Stoelen & Akin as well as Kulik et al., limiting their observations in movements across one line only with no directions or inclinations.	[46, 77, 88]
Q6	The above point in mind, researchers are advised to explore other extensions as well in combined movements. Both Stoelen & Akin's, as well as Kulik's et al. work, extended Shannon's formulation exclusively. Perhaps Hoffmann's or Welford's could be explored/extended as well and their differences reported.	[33, 46, 52, 53, 77, 88, 90]
Q7	Depth perception in virtual environments differs from that of the real world, with users overestimating the distances. Can the work of Machuca & Stuerzlinger, which is to the best of our knowledge the only work that accounted for the depth variable and added to Fitts' (Shannon's) extensions, suffice? With the spike in VR/AR technologies, verification of the above appears to be crucial.	[5, 18, 47, 49, 68, 78, 91]
Q8	Do not evaluate models solely based on the coefficient of determination (R^2). In this work, we explored the limitations of purely relying on the R^2. Instead, aim to include the full line equation (MT) with the reported constants, including standard error and lower as well as upper bound confidence intervals (ideally at CI=95%).	[17, 26]
Q9	Can human factors be accounted for? Tiredness and concentration for example are key elements affecting significantly human performance, yet modelling these factors is challenging and remains a gap in the literature. If this is infeasible, researchers should account for these factors during participant recruitment.	[16, 17, 65, 70, 81, 95]
Q10	Can we reduce the complexities that arise in full 3D space with the use of different sensory modalities? How do other modalities affect object interaction, in particular haptic or auditory feedback? Can we reduce potentially impaired amounts of depth perception by introducing haptic feedback in the sensory interface?	[3, 9, 10, 22, 50, 67, 80–82, 84, 86]
Q11	Manipulation entails different physical influences such as gravity, contact points and the mass of the object. These should prove invaluable, especially in realistic simulations and training scenarios attempting to model performance under realistic settings. Can these properties be modelled? Furthermore, how do different hand types with different contact points and degrees of freedom (e.g. grippers) and their size affect the model?	[13, 46, 81, 82, 88, 96, 97]
Q12	Last but not least, what are the differences in pointing and manipulation, can these two different types of interactions be modelled under one unified formula? More importantly, can all of the aforementioned points be considered and accounted for or should we expect different formulations based on these factors alone?	All of the above including the contents of this work.

Table 2: The table represents potentially important directions and research paths we recommend researchers to pursue for the derivation of a robust performance metric in 3D space. The aforementioned points may prove to be crucial in solidifying such a metric and increase inter-study comparability with the hopes of generalizing it in a multitude of different settings.

There is strong evidence that assessing performance via Fitts' law for 3D pointing tasks appears to suffer primarily due to latency [89], impaired depth perception [6, 51], as well as the lack of haptic (tactile) feedback [9, 22, 50, 80]. The latter two appear to have a direct correlation to what seem to be the benefits of multi-modal interfaces in increasing overall perception [8, 9, 11, 28, 82, 83].

One study investigated 3D virtual hand pointing with the index finger in addition to incorporating vibration feedback using a 3D screen display [63]. While disregarding depth and spatial arrangements, contrary to [5, 59], they found that vibration feedback provides a reasonable addition to visual feedback, which appears to be in line with existing literature on multi-modal interfaces [2, 60, 82]. This is furthermore confirmed by numerous neuroscience studies

indeed confirming that the simultaneous presence of both visual and somatosensory sensory cues is beneficial, particularly due to both modalities overlapping in the same brain region[1, 3, 36, 73].

However, in all cases, sensory conflict can arise when a multi-modal pipeline is unable to stimulate the senses in a synchronised way, which can be counterproductive, resulting in decreased spatial and temporal immersion, effectively nullifying the benefits [66, 69]. While conclusions are indeed difficult to draw as to which modalities directly offer higher quantifiable amounts of perception, we can indirectly measure the human performance and by proposing a standardized metric, inter-study comparisons within the domain of multimodal human-computer interaction may be feasible and increase our understanding.

3.5 Physical Influences

As we mentioned, physical interactions are an integral part of the manipulation process. Firstly, the effects of gravity should be evaluated. To the best of our knowledge, an experimental setting with different levels of gravity has still not been used to assess how it may affect MT and the potential influence it may pose to a model. It still remains a gap to this day [79]. Contact points are another aspect to consider. While there are specific methods and metrics that assess the quality of a given grasp, such as the Largest-minimum resisted wrench [97], it may prove to be beneficial if that can also be accounted for by Fitts' law.

Furthermore, the grasping type users initiate with an object may prove to be a determining factor towards MT and by extent a 3D performance metric covering the manipulation domain. For example, the Southampton Hand Assessment Procedure (SHAP) is a clinically validated hand function test [48] reporting six distinctive grip classifications: precision-tip, lateral, tripod, spherical, power and extension grasping. Vikash Kumar and Emanuel Todorov developed a virtual reality system for hand manipulation based upon a subset of tasks from the SHAP for evaluation purposes [45]. However, the latter did not investigate or include Fitts' model in their evaluation. To which extent these physical properties really affect a model is hard to quantify, yet it should be pursued by researchers that aim to extend a robust performance method towards manipulation.

3.6 Which Input Technology?

Lastly, aggravating a unified model is the multitude of different input devices one can use. As shown by previous work, Fitts' model is heavily influenced by the input technology used, whether that includes tracked wands, a regular mouse, stylus or optical hand tracking [21, 54, 92, 93]. Perhaps proposing a 3D performance model that would aid the community and clear the confusion as to which model is appropriate, may furthermore be aggravated by this factor alone. It should come to no surprise that numerous models should be considered as these may be specific to the input device or accounting for this limitation by adding additional constants which can, however, become problematic [5].

4 CONCLUSION

With the multitude of different performance models published in the HCI community in the last years, confusion still exists as to which formula should be used and rightly so. Yet the collective effort towards a standardized human performance metric in true full 3D space is still missing and work remains largely scattered.

In this review, we analyzed the most prominent and widely used extensions of Fitts' law and their respective contributions and limitations towards the endeavour of deriving a full 3D model. It is not as straightforward as one would think. We observed that not only including all possible spatial arrangements one would expect in full 3D space is challenging to model under one formulation, but also combining translational as well as rotational requirements in tasks is by itself not an easy approach. Furthermore, closing the gap between the discrepancies of pointing and manipulation under potentially one formulation is another very important factor.

We also went beyond the current challenges and also provided a brief outlook of future frontiers that may lay ahead when deriving

a true full 3D human performance model. Factors ranging from how researchers should evaluate their work to the overlooked but yet important human factors may prove to be determining aspects for a "true" 3D performance model.

5 ACKNOWLEDGMENTS

REFERENCES

[1] Asif A. Ghazanfar and Charles E. Schroeder. 2006. Is neocortex essentially multisensory? *Trends in cognitive sciences* 10 (07 2006), 278–85. https://doi.org/10.1016/j.tics.2006.04.008

[2] Jacopo Aleotti, Stefano Bottazzi, and Monica Reggiani. 2002. A Multimodal User Interface for Remote Object Exploration in Teleoperation Systems. https://pdfs.semanticscholar.org/98bd/ce82196ecb7e2f8f0f1a6a480f0a2a00e80d.pdf

[3] Amir Amedi, Rafael Malach, Talma Hendler, Sharon Peled, and Ehud Zohary. 2001. Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience* 4 (2001), 324–330.

[4] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136.

[5] Mayra Donaji Barrera Machuca and Wolfgang Stuerzlinger. 2019. The Effect of Stereo Display Deficiencies on Virtual Hand Pointing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 207, 14 pages. https://doi.org/10.1145/3290605.3300437

[6] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger. 2019. 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)*. IEEE, Osaka, Japan, 585–592. https://doi.org/10.1109/VR.2019.8797975

[7] Xiaojun Bi, Yang Li, and Shumin Zhai. 2013. FFitts Law: Modeling Finger Touch with Fitts' Law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1363–1372. https://doi.org/10.1145/2470654.2466180

[8] Aude Billard and Danica Kragic. 2019. Trends and challenges in robot manipulation. *Science* 364, 6446 (2019), 1149. https://doi.org/10.1126/science.aat8414

[9] D. Brickler, S. V. Babu, J. Bertrand, and A. Bhargava. 2018. Towards Evaluating the Effects of Stereoscopic Viewing and Haptic Interaction on Perception-Action Coordination. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Reutlingen, Germany, 1–516. https://doi.org/10.1109/VR.2018.8446227

[10] Harry Bunt, Robbert-Jan Beun, and Tijn Borghuis. 1998. *Multimodal human-computer communication: systems, techniques, and experiments*. Vol. 1374. Springer Science & Business Media, Berlin-Heidelberg, Germany.

[11] Jennifer L. Burke, Matthew S. Prewett, Ashley A. Gray, Liuquin Yang, Frederick R. B. Stilson, Michael D. Coovert, Linda R. Elliot, and Elizabeth Redden. 2006. Comparing the Effects of Visual-auditory and Visual-tactile Feedback on User Performance: A Meta-analysis. In *Proceedings of the 8th International Conference on Multimodal Interfaces* (Banff, Alberta, Canada) *(ICMI '06)*. ACM, New York, NY, USA, 108–117. https://doi.org/10.1145/1180995.1181017

[12] Rachael A. Burno, Bing Wu, Rina Doherty, Hannah Colett, and Rania Elnaggar. 2015. Applying Fitts' Law to Gesture Based Computer Interactions. *Procedia Manufacturing* 3 (2015), 4342 – 4349. https://doi.org/10.1016/j.promfg.2015.07.429 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.

[13] Jesse Burstyn, Juan Pablo Carrascal, and Roel Vertegaal. 2016. Fitts' Law and the Effects of Input Mapping and Stiffness on Flexible Display Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3649–3658. https://doi.org/10.1145/2858036.2858383

[14] Yeonjoo Cha and Rohae Myung. 2013. Extended Fitts' law for 3D pointing tasks using 3D target arrangements. *International Journal of Industrial Ergonomics* 43, 4 (2013), 350 – 355. https://doi.org/10.1016/j.ergon.2013.05.005

[15] E. R. F. W. Crossman and P. J. Goodeve. 1983. Feedback control of hand-movement and Fitts' law. *The Quarterly Journal of Experimental Psychology Section A* 35, 2 (1983), 251–278. https://doi.org/10.1080/14640748308402133

[16] John de Grosbois, Matthew Heath, and Luc Tremblay. 2015. Augmented feedback influences upper limb reaching movement times but does not explain violations of Fitts' Law. *Frontiers in Psychology* 6 (2015), 800. https://doi.org/10.3389/fpsyg.2015.00800

[17] Heiko Drewes. 2010. Only One Fitts' Law Formula Please!. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI EA '10)*. Association for Computing Machinery, New York, NY, USA, 2813–2822. https://doi.org/10.1145/1753846.1753867

[18] Frank H Durgin, Dennis R Proffitt, Thomas J Olson, and Karen S Reinke. 1995. Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance* 21, 3 (1995), 679.

[19] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[20] Paul M Fitts and James R Peterson. 1964. Information capacity of discrete motor responses. *Journal of experimental psychology* 67, 2 (1964), 103.

[21] Clifton Forlines and Ravin Balakrishnan. 2008. 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* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1563–1572. https://doi.org/10.1145/1357054.1357299

[22] F. Gemperle, N. Ota, and D. Siewiorek. 2001. Design of a wearable tactile display. In *Proceedings Fifth International Symposium on Wearable Computers.* IEEE, Zurich, Switzerland, 5–12. https://doi.org/10.1109/ISWC.2001.962082

[23] Douglas J. Gillan, Kritina Holden, Susan Adam, Marianne Rudisill, and Laura Magee. 1990. How Does Fitts' Law Fit Pointing and Dragging?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 227–234. https://doi.org/10.1145/97243.97278

[24] Douglas J Gillan, Kritina Holden, Susan Adam, Marianne Rudisill, and Laura Magee. 1992. How should fitts' law be applied to human-computer interaction? *Interacting with Computers* 4, 3 (1992), 291 – 313. https://doi.org/10.1016/0953-5438(92)90019-C

[25] A. Godse, R. Khadka, and A. Banic. 2019. Evaluation of Visual Perception Manipulation in Virtual Reality Training Environments to Improve Golf Performance. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 1807–1812.

[26] Julien Gori, Olivier Rioul, Yves Guiard, and Michel Beaudouin-Lafon. 2018. The Perils of Confounding Factors: How Fitts' Law Experiments Can Lead to False Conclusions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3173770

[27] Tovi Grossman and Ravin Balakrishnan. 2004. Pointing at Trivariate Targets in 3D Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 447–454. https://doi.org/10.1145/985692.985749

[28] Rakesh Gupta, Daniel Whitney, and David Zeltzer. 1997. Prototyping and design for assembly analysis using multimodal virtual environments. *Computer-Aided Design* 29, 8 (1997), 585–597.

[29] T. Ha and W. Woo. 2010. An empirical evaluation of virtual hand techniques for 3D object manipulation in a tangible augmented reality environment. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, Waltham, MA, USA, 91–98. https://doi.org/10.1109/3DUI.2010.5444713

[30] Robert F. Hess, Long To, Jiawei Zhou, Guangyu Wang, and Jeremy R. Cooperstock. 2015. Stereo Vision: The Haves and Have-Nots. *i-Perception* 6, 3 (2015), 2041669515593028. https://doi.org/10.1177/2041669515593028

[31] Errol R. Hoffmann. 1991. A comparison of hand and foot movement times. *Ergonomics* 34, 4 (1991), 397–406. https://doi.org/10.1080/00140139108967324 PMID: 1860460.

[32] Errol R. Hoffmann. 1995. Effective target tolerance in an inverted Fitts task. *Ergonomics* 38, 4 (1995), 828–836. https://doi.org/10.1080/00140139508925153

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

[34] D. E. Holmes, D. K. Charles, P. J. Morrow, S. McClean, and S. M. McDonough. 2016. Using Fitt's Law to Model Arm Motion Tracked in 3D by a Leap Motion Controller for Virtual Reality Upper Arm Stroke Rehabilitation. In *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, Dublin, Ireland, 335–336.

[35] Hyungki Hong and Seok Hyon Kang. 2015. Measurement of the lens accommodation in viewing stereoscopic displays. *Journal of the Society for Information Display* 23, 1 (2015), 19–26. https://doi.org/10.1002/jsid.303

[36] Thomas James, G Keith Humphrey, Sabiha Gati, Philip Servos, Ravi S Menon, and Melvyn Goodale. 2002. Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* 40 (02 2002), 1706–14. https://doi.org/10.1016/S0028-3932(02)00017-9

[37] Izabelle Janzen, Vasanth K. Rajendran, and Kellogg S. Booth. 2016. Modeling the Impact of Depth on Pointing Performance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 188–199. https://doi.org/10.1145/2858036.2858244

[38] Tiffany S Jastrzembski and Neil Charness. 2007. The Model Human Processor and the older adult: Parameter estimation and validation within a mobile phone task. *Journal of experimental psychology: applied* 13, 4 (2007), 224.

[39] Robert V Kenyon and Stephen R Ellis. 2014. *Vision, perception, and object manipulation in virtual environments.* Springer, New York, USA. 47–70 pages.

[40] Robert Kerr. 1973. Movement Time in an Underwater Environment. *Journal of Motor Behavior* 5, 3 (1973), 175–178. https://doi.org/10.1080/00222895.1973.10734962 PMID: 23961747.

[41] A. A. KNIGHT and P. R. DAGNALL. 1967. Precision in Movements. *Ergonomics* 10, 3 (1967), 321–330. https://doi.org/10.1080/00140136708930874 PMID: 6077520.

[42] G. V. Kondraske. 1994. 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. IEEE, Baltimore, MD, USA, 307–308 vol.1.

[43] Georgios Kouroupetroglou, Alexandros Pino, Athanasios Balmpakakis, Dimitrios Chalastanis, Vasileios Golematis, Nikolaos Ioannou, and Ioannis Koutsoumpas. 2012. Using Wiimote for 2D and 3D Pointing Tasks: Gesture Performance Evaluation. In *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*. Springer Berlin Heidelberg, Berlin, Heidelberg, 13–23.

[44] A. Kulik, A. Kunert, and B. Froehlich. 2020. On Motor Performance in Virtual 3D Object Manipulation. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 2041–2050.

[45] V. Kumar and E. Todorov. 2015. MuJoCo HAPTIX: A virtual reality system for hand manipulation. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, Seoul, South Korea, 657–663. https://doi.org/10.1109/HUMANOIDS.2015.7363441

[46] André Kunert, Alexander Kulik, Anke Huckauf, and Bernd Fröhlich. 2007. 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* (Weimar, Germany) *(EGVE'07)*. Eurographics Association, Goslar, DEU, 43–52.

[47] Donald R. Lampton, Daniel P. McDonald, Michael Singer, and James P. Bliss. 1995. Distance Estimation in Virtual Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 39, 20 (1995), 1268–1272. https://doi.org/10.1177/154193129503902006

[48] Colin M. Light, Paul H. Chappell, and Peter J. Kyberd. 2002. Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: Normative data, reliability, and validity. *Archives of Physical Medicine and Rehabilitation* 83, 6 (2002), 776–783. https://doi.org/10.1053/apmr.2002.32737

[49] Chiuhsiang Joe Lin and Bereket Haile Woldegiorgis. 2015. Interaction and visual performance in stereoscopic displays: A review. *Journal of the Society for Information Display* 23, 7 (2015), 319–332.

[50] L. Liu, R. van Liere, C. Nieuwenhuizen, and J. Martens. 2009. Comparing Aimed Movements in the Real World and in Virtual Reality. In *2009 IEEE Virtual Reality Conference*. IEEE, Lafayette, LA, USA, 219–222.

[51] P. Lubos, G. Bruder, and F. Steinicke. 2014. Analysis of direct selection in head-mounted display environments. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, Minneapolis, MN, USA, 11–18.

[52] I. Scott MacKenzie. 1992. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Hum.-Comput. Interact.* 7, 1 (March 1992), 91–139. https://doi.org/10.1207/s15327051hci0701_3

[53] I. Scott MacKenzie and Poika Isokoski. 2008. Fitts' Throughput and the Speed-Accuracy Tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1633–1636. https://doi.org/10.1145/1357054.1357308

[54] I. Scott MacKenzie, Abigail Sellen, and William A. S. Buxton. 1991. A Comparison of Input Devices in Element Pointing and Dragging Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New Orleans, Louisiana, USA) *(CHI '91)*. Association for Computing Machinery, New York, NY, USA, 161–166. https://doi.org/10.1145/108844.108868

[55] Mike McGee, Brian Amento, Patrick Brooks, and Hope Harley. 1997. Fitts and VR: Evaluating Display and Input Devices with Fitts' Law. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 41, 2 (1997), 1259–1262. https://doi.org/10.1177/1071181397041002119

[56] Sean A. McGlynn and Wendy A. Rogers. 2017. Considerations for Presence in Teleoperation. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. Association for Computing Machinery, New York, NY, USA, 203–204. https://doi.org/10.1145/3029798.3038369

[57] M. Alejandra Menchaca-Brandan, Andrew M. Liu, Charles M. Oman, and Alan Natapoff. 2007. Influence of Perspective-Taking and Mental Rotation Abilities in Space Teleoperation. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (Arlington, Virginia, USA) *(HRI '07)*. Association for Computing Machinery, New York, NY, USA, 271–278. https://doi.org/10.1145/1228716.1228753

[58] David E Meyer, Richard A Abrams, Sylvan Kornblum, Charles E Wright, and JE Keith Smith. 1988. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological review* 95, 3 (1988), 340.

[59] Atsuo Murata and Hirokazu Iwase. 2001. Extending Fitts' law to a three-dimensional pointing task. *Human movement science* 20, 6 (2001), 791–805.

[60] Anne M. Murray, Roberta L. Klatzky, and Pradeep K. Khosla. 2003. Psychophysical Characterization and Testbed Validation of a Wearable Vibrotactile Glove for Telemanipulation. *Presence: Teleoper. Virtual Environ.* 12, 2 (April 2003), 156–182. https://doi.org/10.1162/105474603321640923

[61] Michael Narayan, Leo Waugh, Xiaoyu Zhang, Pradyut Bafna, and Doug Bowman. 2005. Quantifying the Benefits of Immersion for Collaboration in Virtual Environments. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (Monterey, CA, USA) *(VRST '05)*. Association for Computing Machinery, New York, NY, USA, 78–81. https://doi.org/10.1145/1101616.1101632

[62] K. Nieuwenhuizen, L. Liu, R. v. Liere, and J. Martens. 2009. Insights from Dividing 3D Goal-Directed Movements into Meaningful Phases. *IEEE Computer Graphics and Applications* 29, 6 (2009), 44–53.

[63] M. Pfeiffer and W. Stuerzlinger. 2015. 3D virtual hand pointing with EMS and vibration feedback. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, Arles, France, 117–120.

[64] Alexandros Pino, Evangelos Tzemis, Nikolaos Ioannou, and Georgios Kouroupetroglou. 2013. Using Kinect for 2D and 3D Pointing Tasks: Performance Evaluation. In *Proceedings of the 15th International Conference on Human-Computer Interaction: Interaction Modalities and Techniques - Volume Part IV* (Las Vegas, NV) *(HCI'13)*. Springer-Verlag, Berlin, Heidelberg, 358–367. https://doi.org/10.1007/978-3-642-39330-3_38

[65] Céline Poletti, Rita Sleimen-Malkoun, Leslie Marion Decker, Frédérique Retornaz, Patrick Lemaire, and Jean-Jacques Temprado. 2017. Strategic variations in Fitts' task: comparison of healthy older adults and cognitively impaired patients. *Frontiers in aging neuroscience* 8 (2017), 334.

[66] George V. Popescu, Grigore C. Burdea, and Helmuth Trefftz. 2002. *Multimodal interaction modeling*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 435–454.

[67] Francis Quek, David McNeill, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCullough, and Rashid Ansari. 2002. Multimodal Human Discourse: Gesture and Speech. *ACM Trans. Comput.-Hum. Interact.* 9, 3 (Sept. 2002), 171–193. https://doi.org/10.1145/568513.568514

[68] Rebekka S. Renner, Boris M. Velichkovsky, and Jens R. Helmert. 2013. The Perception of Egocentric Distances in Virtual Environments - A Review. *ACM Comput. Surv.* 46, 2, Article 23 (Dec. 2013), 40 pages. https://doi.org/10.1145/2543581.2543590

[69] Paul Richard, Grigore Burdea, Daniel Gomez, and Philippe Coiffet. 1994. A comparison of haptic, visual and auditive force feedback for deformable virtual objects.

[70] V. Rozand, F. Lebon, C. Papaxanthis, and R. Lepers. 2015. Effect of mental fatigue on speed–accuracy trade-off. *Neuroscience* 297 (2015), 219 – 230. https://doi.org/10.1016/j.neuroscience.2015.03.066

[71] Ana Sacau, Jari Laarni, and Tilo Hartmann. 2008. Influence of individual factors on presence. *Computers in Human Behavior* 24, 5 (2008), 2255–2273. https://doi.org/10.1016/j.chb.2007.11.001

[72] Farzan Sasangohar, I. Scott MacKenzie, and Stacey D. Scott. 2009. Evaluation of Mouse and Touch Input for a Tabletop Display Using Fitts' Reciprocal Tapping Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 53, 12 (2009), 839–843. https://doi.org/10.1177/154193120905301216

[73] K Sathian and A Zangaladze. 2002. Feeling with the mind's eye: Contribution of visual cortex to tactile perception. *Behavioural brain research* 135 (10 2002), 127–32. https://doi.org/10.1016/S0166-4328(02)00141-9

[74] W. N. Schofield. 1976. Do Children Find Movements Which Cross the Body Midline Difficult? *Quarterly Journal of Experimental Psychology* 28, 4 (1976), 571–582. https://doi.org/10.1080/14640747608400584

[75] Garth Shoemaker, Takayuki Tsukitani, Yoshifumi Kitamura, and Kellogg S. Booth. 2012. Two-Part Models Capture the Impact of Gain on Pointing Performance. *ACM Trans. Comput.-Hum. Interact.* 19, 4, Article 28 (Dec. 2012), 34 pages. https://doi.org/10.1145/2395131.2395135

[76] Richard HY So and Michael J Griffin. 2000. Effects of a target movement direction cue on head-tracking performance. *Ergonomics* 43, 3 (2000), 360–376.

[77] Martin F. Stoelen and David L. Akin. 2010. Assessment of Fitts' Law for Quantifying Combined Rotational and Translational Movements. *Human Factors* 52, 1 (2010), 63–77. https://doi.org/10.1177/0018720810366560 PMID: 20653226.

[78] J. E. Swan, G. Singh, and S. R. Ellis. 2015. Matching and Reaching Depth Judgments with Real and Augmented Reality Targets. *IEEE Transactions on Visualization and Computer Graphics* 21, 11 (Nov 2015), 1289–1298. https://doi.org/10.1109/TVCG.2015.2459895

[79] Robert J. Teather and Wolfgang Stuerzlinger. 2007. Guidelines for 3D Positioning Techniques. In *Proceedings of the 2007 Conference on Future Play* (Toronto, Canada) *(Future Play '07)*. Association for Computing Machinery, New York, NY, USA, 61–68. https://doi.org/10.1145/1328202.1328214

[80] Robert J. Teather and Wolfgang Stuerzlinger. 2011. Pointing at 3D Targets in a Stereo Head-Tracked Virtual Environment. In *Proceedings of the 2011 IEEE Symposium on 3D User Interfaces (3DUI '11)*. IEEE Computer Society, USA, 87–94.

[81] Zachary C. Thumser, Andrew B. Slifkin, Dylan T. Beckler, and Paul D. Marasco. 2018. Fitts' Law in the Control of Isometric Grip Force With Naturalistic Targets. *Frontiers in Psychology* 9 (2018), 560. https://doi.org/10.3389/fpsyg.2018.00560

[82] E. Triantafyllidis, C. Mcgreavy, J. Gu, and Z. Li. 2020. Study of Multimodal Interfaces and the Improvements on Teleoperation. *IEEE Access* 8 (2020), 78213–78227. https://doi.org/10.1109/ACCESS.2020.2990080

[83] Eleftherios Triantafyllidis, Chuanyu Yang, Christopher McGreavy, Wenbin Hu, and Zhibin Li. 2020. *Robot intelligence for real-world applications*. IET Computing and Networks, UK. 63 pages.

[84] Matthew Turk. 2014. Multimodal interaction: A review. *Pattern Recognition Letters* 36 (2014), 189 – 195. https://doi.org/10.1016/j.patrec.2013.07.003

[85] Jonathan Vaughan, Deborah A Barany, Anthony W Sali, Steven A Jax, and David A Rosenbaum. 2010. Extending Fitts' Law to three-dimensional obstacle-avoidance movements: support for the posture-based motion planning model. *Experimental brain research* 207, 1-2 (2010), 133–138.

[86] Roel Vertegaal. 2008. A Fitts Law Comparison of Eye Tracking and Manual Input in the Selection of Visual Targets. In *Proceedings of the 10th International Conference on Multimodal Interfaces* (Chania, Crete, Greece) *(ICMI '08)*. Association for Computing Machinery, New York, NY, USA, 241–248. https://doi.org/10.1145/1452392.1452443

[87] Yanqing Wang and Christine L. MacKenzie. 1999. Object Manipulation in Virtual Environments: Relative Size Matters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 48–55. https://doi.org/10.1145/302979.302989

[88] Yanqing Wang, Christine L. MacKenzie, Valerie A. Summers, and Kellogg S. Booth. 1998. The Structure of Object Transportation and Orientation in Human-Computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA) *(CHI '98)*. ACM Press/Addison-Wesley Publishing Co., USA, 312–319. https://doi.org/10.1145/274644.274688

[89] Colin Ware and Ravin Balakrishnan. 1994. Reaching for Objects in VR Displays: Lag and Frame Rate. *ACM Trans. Comput.-Hum. Interact.* 1, 4 (Dec. 1994), 331–356. https://doi.org/10.1145/198425.198426

[90] A. T. (Alan Traviss) Welford. 1968. *Fundamentals of skill.* Methuen, London.

[91] Bob G. Witmer and Paul B. Kline. 1998. Judging Perceived and Traversed Distance in Virtual Environments. *Presence: Teleoper. Virtual Environ.* 7, 2 (April 1998), 144–167. https://doi.org/10.1162/105474698565640

[92] Jacob O. Wobbrock, Edward Cutrell, Susumu Harada, and I. Scott MacKenzie. 2008. An Error Model for Pointing Based on Fitts' Law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1613–1622. https://doi.org/10.1145/1357054.1357306

[93] Jacob O. Wobbrock, Alex Jansen, and Kristen Shinohara. 2011. Modeling and Predicting Pointing Errors in Two Dimensions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1653–1656. https://doi.org/10.1145/1978942.1979183

[94] Charles M. Zaroff, Magosha Knutelska, and Thomas E. Frumkes. 2003. Variation in Stereoacuity: Normative Description, Fixation Disparity, and the Roles of Aging and Gender. *Investigative Ophthalmology & Visual Science* 44, 2 (02 2003), 891–900. https://doi.org/10.1167/iovs.02-0361

[95] Xiaolu Zeng, Alan Hedge, and Francois Guimbretiere. 2012. Fitts' Law in 3D Space with Coordinated Hand Movements. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56, 1 (2012), 990–994. https://doi.org/10.1177/1071181312561207

[96] Shumin Zhai, Paul Milgram, and William Buxton. 1996. 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* (Vancouver, British Columbia, Canada) *(CHI '96)*. Association for Computing Machinery, New York, NY, USA, 308–315. https://doi.org/10.1145/238386.238534

[97] Y. Zheng. 2013. An Efficient Algorithm for a Grasp Quality Measure. *IEEE Transactions on Robotics* 29, 2 (2013), 579–585.