Positional Variance Profiles (PVPs): A New Take on the Speed-Accuracy Trade-off

Julien Gori, Quentin Bellut

Positional Variance Profiles (PVPs): A New Take on the Speed-Accuracy Trade-off

JULIEN GORI, ISIR, CNRS, Sorbonne Université, France
QUENTIN BELLUT, Sorbonne Université, France

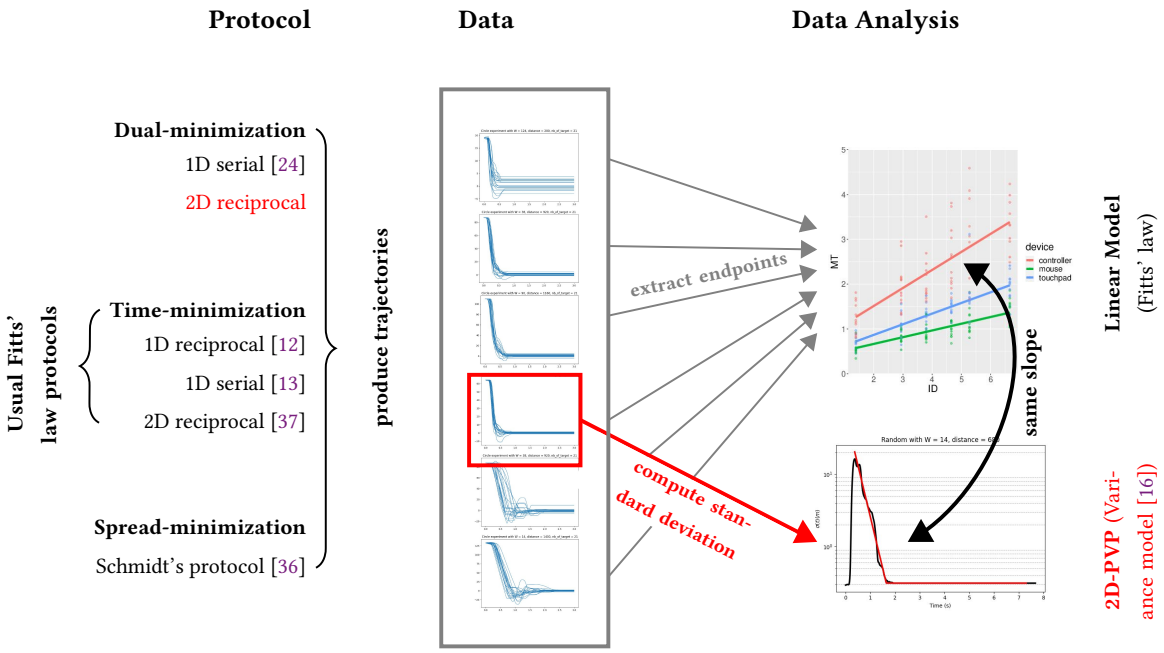

Fig. 1. In this work (red), we transpose our behavior model [16] called the Variance Model, to propose a novel 2D random *dual-minimization* protocol, as well as the 2D PVP method, which extracts features from the standard deviation of a set of trajectories. When evaluating human performance, this provides an alternative to classical Fitts' law experiments, which use *time-minimization* protocols, and endpoint information only. PVPs can be computed from a single experimental condition (red rectangle), whereas estimating Fitts' law typically requires multiple conditions (gray rectangle); Fitts' law parameters can also be computed from PVP features. This is possible because PVPs exploit information from the whole trajectory, and not just endpoints as in Fitts' law.

Fitts' law is a behavioral model, used to design protocols and analyze data from pointing experiments. These are usually conducted in HCI to evaluate input performance. We recently proposed an alternative method to characterize input performance, called the method of PVPs in 1D, based on 1) a dual-minimization protocol, and 2) an analysis of the variability of entire trajectories. We extend the method in 2D; our contributions include new metrics, a new protocol, and a Python library. We also present the results of a controlled experiment where the new method is validated using three devices (mouse, touchpad, controller): effect sizes in the 2D case replicate those previously found. We also propose a comparison between Fitts' law and our novel evaluation: the method of PVPs provides more information than Fitts' law, and can predict its parameters. We discuss how this new method may relieve open problems of Fitts' law.

CCS Concepts: • **Human-centered computing** → **HCI theory, concepts and models**.

Additional Key Words and Phrases: Fitts' law, PVP, pointing, evaluation

ACM Reference Format:
Julien Gori and Quentin Bellut. 2023. Positional Variance Profiles (PVPs): A New Take on the Speed-Accuracy Trade-off. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3544548.3581071

1 INTRODUCTION

In human motor control, an increase in speed is invariably made at the cost of a decrease in accuracy — a speed-accuracy trade-off that has been captured by several behavioral models. One of the most popular in HCI is Fitts' law; it predicts the time MT it takes a participant to reach a target of size W, located at a distance D as fast as possible.

$$MT = a + b \log_2\left(1 + \frac{D}{W}\right) \equiv a + b\, ID. \tag{1}$$

Here, ID is expressed in bits, because of the base 2 logarithm. In the remainder of this paper, all logarithms will be base 2, although we omit the subscript. Ever since the seminal study by Card *et al.* [4] which used Fitts' law as a basis to evaluate input performance, it has become a staple of HCI research: in so-called Fitts' law experiments, it serves as a reference behavioral model, and is used to design experimental protocols and fit data when researchers evaluate input performance [18, 37] (see Fig. 1)

However:

(1) There is little knowledge on why Fitts' law is a good description of human behavior. The theoretical basis of the law is not well understood: the experimental protocol and data analysis are based on Fitts' loose analogy between an aiming task and a noisy channel [12, 37]. That analogy however has been critiqued multiple times [7, 18, 30], which makes some Fitts' law practices questionable.
(2) Fitts' law experiments use time-minimization protocols, which poorly control the accuracy of movements [24, 25]: movements may end up outside the target —usually for small targets— or be clustered within a small part of the target —usually for large targets [12, 37].[1] In practice, the independent variable W is often corrected *post-hoc* to reflect the actual spread of endpoints via the so-called effective version of Fitts' law [37], a practice that is contrary to the purpose of a controlled experiment —the independent variable should not be adjusted after the experiment— and has further been shown to be arbitrary [18].

As a result, many problems around Fitts' law experiments remain unresolved: Several versions of ID in Eq. (1) exist [18, 34, 37]. Which one should we use and why? What is the interpretation of the intercept a in Eq. (1)? Should it be equal to 0 [23, 37] as prescribed by the original information-theoretic description [12, 19]? Interpreting the intercept is important because it conditions whether an assessment of input performance should consider the intercept or not, and whether a non-zero value is indicative of a poor experiment or not as suggested by Soukoreff and MacKenzie [37]. Can throughput, a one-dimensional score that estimates the "performance" of a device or technique, be unequivocally defined, and what then is its interpretation [18, 37, 41]? Should the experimenter encourage the participants to aim for a 4% miss rate as often encouraged [18, 37], or should the experimenter strive for a null miss rate?

In this work, we build on our recent information-theoretic transmission model with feedback, which we call the *Variance Model* [16], and the corresponding data analysis method called the *method of PVPs* (Positional Variance Profiles) to provide an alternative to Fitts' law experiments for input performance evaluations. The method of PVPs is a data aggregation method, where a time series of the variance of a *set* of trajectories is computed to obtain a variance profile.

[1] Fitts himself was aware of this. His 1954 paper [12] features 3 experiments: The first one is the one who was popularized, but the latter two featured variations where the width was properly controlled, by means of pegs and discs.

PVPs display two phases; we described the second phase with a theoretical model, and the first phase empirically only [16].

The method of PVPs can be applied to several protocols including ones where there is no pre-specified width such as dual minimization protocols and is also more expressive than Fitts' law, because it characterizes whole trajectories, and not just endpoints (see Fig. 1). Another nice property of PVPs is that Fitts' law can essentially be derived from them. Hence, by transposing the method of PVPs in HCI, we also alleviate some drawbacks associated with Fitts' law:

- Fitts' law is expressed as the interaction between the two phases of the PVP. As a result, Fitts' law parameters can be computed from the PVP parameters and can be interpreted within the information-transmission scheme. This provides new arguments to the aforementioned unresolved problems with Fitts' law experiments.
- We propose a new protocol, amenable to the method of PVPs, which does not try to enforce endpoint width constraints to generate data, thereby removing the need for a *post-hoc* correction. The protocol is also simpler to conduct, since it does not require crossing D and W: a single experimental condition is enough.

Currently, however, the method of PVPs can not be applied off-the-shelf to most HCI evaluation opportunities: the method was described for 1D data only, whereas most situations in HCI feature two-dimensional pointing. A similar problem affected Fitts' law experiments, which was also initially described in 1D only, and which attracted significant effort [1, 21, 28, 31, 39]. Thus, in this work we extend the one dimensional method of PVP to so-called 2D-PVPs. In doing so, we make the connection to existing literature on 2D Fitts' law whenever that is possible. Further, the effect of D on the PVP features was only assessed on 2 levels of D in [16]. In this work, we submit the method of PVPs to a tougher examination by using 6 levels of D. We also present the first PVP-based evaluation of multiple devices (mouse, controller, touchpad), and its comparison with Fitts'-law based evaluations. In summary, our contributions are:

- transposing a theory and method that describes the evolution of a set of pointing trajectories to HCI ;
- proposing a new dual-minimization protocol for evaluating input performance, that does not require crossing the factors D and W as in Fitts' law experiments;
- extending the method of PVPs to the 2D case, proposing 3 different profiles;
- estimating the effect of D on PVP features for multiple levels of D, further substantiating the empirical analysis of the first phase;
- contrasting a classical Fitts' law experiment with our new experiment and method when comparing three devices (mouse, touchpad, controller);
- providing new arguments to open problems surrounding Fitts' law experiments.
- A Python package called *PVPlib* that implements the PVP methods presented in this paper.

2 BACKGROUND

2.1 Fitts' law and Fitts' law experiments

The nominal version [42] of Fitts' law[2] is given Eq. (1). Fitts' law in its *effective* or *corrected* version [5, 30] reads

$$MT = a + b \log\left(1 + \frac{D}{4.133\sigma}\right) \equiv a + b\,\mathrm{ID}_e, \tag{2}$$

where σ is the standard deviation of endpoints measured for each condition. This is a popular amendment to Fitts' nominal formula, and is meant to account for divergent participant speed-accuracy strategies [5, 18, 37], as some

[2]This is the MacKenzie "+1" version. Fitts' original version is also popular, see [19].

participants emphasize speed but miss the target more often in the process. As explained in the introduction and by Guiard [24, 25], such a post-hoc adjustment is not in line with controlled experiments, where it is assumed that the independent variable is controlled by the experimenter, and not up to the participant or the environment.

When Fitts' law is used to assess the performance of a participant or a device —*i.e.*, a Fitts' law experiment is performed— one usually runs a controlled experiment where the factors D and W are fully crossed. The law's parameters, a and b in Eqs. (1) and (2), are then empirically estimated from endpoint information (time and position), see Fig. 1. Overall, lower values of a (intercept) and b (slope) indicate a better efficiency since they lead to shorter movement times. A one dimensional score known as the *ISO-throughput* [37]

$$\mathrm{TP} = \frac{1}{N}\sum_{n=1}^{N}\frac{\mathrm{ID}_{e,n}}{\mathrm{MT}_n}, \tag{3}$$

is sometimes used to condense all measured input performance into a single score, where N is the number of different experimental conditions, and where the data are aggregated by experimental condition. That metric however has been criticized on practical and theoretical grounds [18, 41].

2.2 Evaluating Input Performance: Protocols

Fitts' law experiments are part of a wider family of controlled studies that are used to investigate the speed-accuracy trade-off. These experiments may use different protocols to produce data, of which Guiard [25] distinguishes three classes (see Fig. 1):

- time-minimization protocols, where participants are instructed to hit targets as fast as possible. Most protocols used in Fitts' law experiments, such as Fitts' reciprocal paradigm [12], Fitts' serial paradigm [13], and ISO's multi-directional tapping task [37] are time-minimization protocols;
- spread-minimization protocols, where participants are instructed to hit targets under a movement time constraint while minimizing the spread of endpoints. One well known spread-minimization protocol is due to Schmidt [36], and is associated with an empirical law known as Schmidt's law, but it has not been very popular in HCI;
- dual-minimization protocols, where participants should minimize both pointing time and spread of endpoints. For these protocols, the experimenter may direct the participant strategy (*e.g.*, emphasize speed over accuracy). Guiard and colleagues [24, 25] previously introduced a 1D serial dual-minimization protocol.

Contrary to Fitts' law, which requires a target width to be predetermined, and thus a time-minimization protocol to be evaluated, the method of PVPs can be applied irrespective of the protocol, since it is based only on computing a time series of the variance of a set of trajectories, and does not require a predetermined target width. In our first work on the variance model [16], we tested the method of PVPs on a time-minimization paradigm, and a dual minimization paradigm, with similar outcomes (see Properties in the next section). In this work, we propose a novel 2D dual minimization protocol where the direction of movements are randomized, also with similar outcomes.

3 METHOD OF PVPS

In this section, we review how Positional Variance Profiles (PVPs) are constructed, how they are fitted to extract key features, and their connection to an information-transmission scheme [16] that explains human behavior as the result of optimal transmission over a noisy Gaussian channel with feedback information.

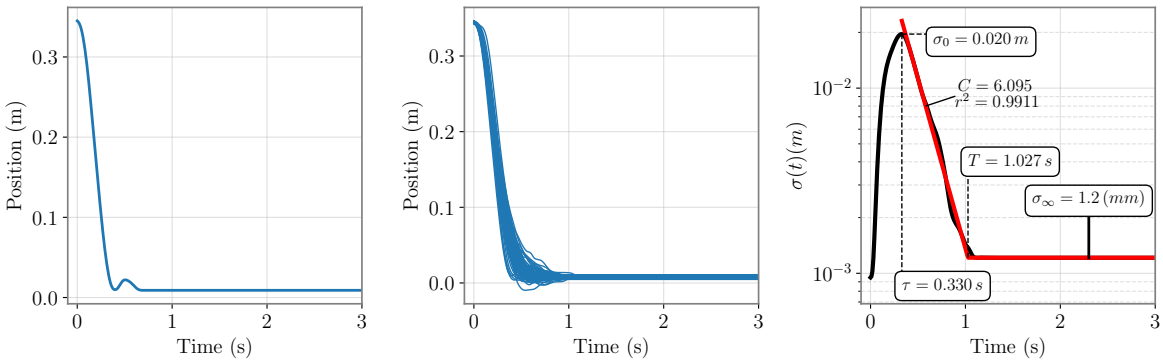

Fig. 2. PVP construction, illustrated on trajectories from a block of the Pointing Dynamics Dataset [32] (P1, D = 0.353, ID = 6). Left panel: An individual trajectory, with position plotted against time. Middle panel: A set of pooled, synchronized trajectories, with position plotted against time. Right panel: PVP of the set of movements used in the middle panel, where standard deviation of the position plotted against time is plotted in black in a lin-log scale. The spline fit is displayed in red, as well as meaningful features of that fit: τ the duration of the first phase, σ_0 the initial variance at the start of the second phase, σ_∞ the end variance at the end of the second phase, C the estimated rate of the second phase, and T, the effective movement time.

3.1 Constructing PVPs

A trajectory is a time series of coordinates *e.g.*, mouse cursor or fingertip positions, for a single movement. PVPs are computed for all trajectories of a given block by:

(1) Extending each trajectory so that all movements have the same duration, by *padding* the final value of position (to, say, 3 s), and resampling them to have equally spaced samples (Left panel in Fig. 2). By extending, we do not mean warping the trajectory as in *e.g.*, Dynamic Time Warping [35]; instead we do as if the recording lasted a little longer while movements had already ended.
(2) Synchronizing all trajectories, using their starting time as the new time origin (Middle panel in Fig. 2).
(3) Computing the standard deviation[3] $\sigma(t)$ of the position signal over time (Right panel, black curve in Fig. 2). The profile is best represented in a lin-log scale (see Property 6 below).

3.2 Properties of PVPs

Property 1. *PVPs can be decomposed into three phases: A first phase where positional variance increases rapidly, a second phase where it decreases steadily, and a third phase where positional variance is constant.*

Note that while [16] discusses a two-phase model, PVPs actually have three phases: this is because the third phase is due to how trajectories are extended. It does not have any meaning beyond the fact it is needed to fit the PVPs correctly.

The features of the first phase are (see Right panel of Fig. 2)

- τ, the duration of the first phase;
- D_τ, the distance covered during the first phase (not shown);
- σ_0, the standard deviation of position at the end of the first phase (*i.e.*, start of the second phase).

[3]While the name PVP suggests we deal with variance, the PVP actually is a time series of standard deviations. The difference is slight: in log-scale the two are related by a factor 2, and it turns out to be more convenient to deal with standard deviations in the theoretical model due to the 2 factor cancelling out.

The first phase of the PVPs was not described by a theoretical model, but was instead empirically characterized with linear models [16]:

Property 2. *The duration of the first phase, τ, is linear in D:*

$$\tau = \tau_0 + \alpha D \tag{4}$$

Property 3. *The distance covered during the first phase, D_τ, is proportional to D*

$$D_\tau = \beta D \tag{5}$$

Property 4. *The standard deviation at the end of the first phase (maximum standard deviation), σ_0, is proportional to D*

$$\sigma_0 = \gamma D \tag{6}$$

Deviations from the average trajectory in the second phase were also characterized:

Property 5. *Deviations from the average trajectory are Gaussian distributed.*

In [16], we made the link between the second phase and a biologically plausible information-transmission model. We predicted that for a Gaussian input, a feedforward channel with Gaussian noise, and a feedback information channel with delay, the standard deviation of the output can decrease exponentially over time *at best*, at a rate C called the *capacity*, which is a participant-dependent constant, independent of task parameters such as D and W.

Property 6. *During the second phase, PVPs can decrease at best at an exponential rate C that is independent of the task properties:*
In the second phase

$$\sigma(t) = \sigma_0 2^{-C(t-\tau)}, \text{ or equivalently} \tag{7}$$

$$\log \sigma(t + \Delta t) = \log \sigma(t) - C\Delta t. \tag{8}$$

This property results from an optimization procedure that considers a set of trajectories at once, and that is agnostic of the task— if the controller (in this case, the human) behaves optimally, has feedback information of the location of the endpoint they are controlling and has their actions subjected to Gaussian noise, then this is the best they could achieve. Users could perform suboptimally for many reasons *e.g.*, poor feedback information, but could only do better if assumptions the model makes are violated *e.g.*, if they had access to another source of information or if they were not subjected to Gaussian noise.

It may come as a surprise that the model, through Prop. 6 places a limit on the rate at which the standard deviation of position decreases, and not, say, velocity or acceleration. The assumption is that the limiting factor when close to the target is information processing due to noise in the perceptuomotor loop, more so than dynamical considerations such as limb inertia or limits on torques.

3.3 Fitting PVPs

PVPs are fitted based on Properties 1 and 6. The standard deviation profile is logarithmically transformed to simplify the estimation of C; a linear spline is then fitted to the second and third phases. A least-squares method determines the knot of the spline (T in the right panel of Fig 2), as well as the slope of the first part of the spine (C in the right panel of Fig. 2); the slope of the third phase is assumed null, movements having terminated. This method uses the

entire trajectory in the second and third phase to estimate T, which makes it more robust than alternatives based on differentiating the PVP.

3.4 Link between PVPs and Fitts' law parameters

The empirical model of the first phase (Props. (2–4)) interacts with the variance decreasing model of the second phase (Prop. 6), through σ_0: the first phase "generates" σ_0, which can only be reduced at a fixed rate during the second phase. The task's parameters influence the two phases: a larger D will lead to higher speeds early in the movement, and thus larger σ_0 [36] while a smaller W will lead to a smaller endpoint variance. Both these cases will thus tend to increase the duration of the second phase. Eqs. 27 and 29 in [16], obtained by summing the duration of both phases, provide the estimated average movement durations T:

$$\text{Effective version:} \quad T = \tau'(\gamma) + 1/C \ \log\left(\frac{D}{4.133\sigma_\infty}\right), \tag{9}$$

$$\text{Nominal version:} \quad T = \tau''(\gamma, \varepsilon) + 1/C \ \log\left(\frac{D}{W}\right), \tag{10}$$

where ε is the miss rate and γ is defined in Eq. (6). The exact nature of the intercepts τ' and τ" is discussed further in Sect. 8.3.

Identification. Comparing Eqs. (9) and (10) with Eqs. (2) and (1), it is straightforward to identify:

(1) T with MT, where T acts as a proxy for MT [16, Fig. 7],
(2) C with $1/b$, the inverse of the slope in Fitts' law,
(3) σ_∞ with the endpoint standard deviation σ,
(4) τ' and τ'' with a, the intercept in Fitts' law.

Thus, Fitts' law parameters can be predicted by PVP features and interpreted within the information-transmission model; We previously found these predictions to be accurate for a mouse in the 1D case [16]. In Sect. 6, we evaluate the reliability of that prediction in the 2D case with different devices.

4 EXTENDING THE METHOD OF PVPS IN 2D

In most HCI applications, movements are performed in two dimensions *e.g.*, moving a mouse cursor on a computer screen, which means we ought to extend the method of PVP in 2D. Instead of proposing a multidimensional information-transmission model, we look for a 1D proxy of the 2D spread of data. Note that 2D versions of Fitts' law were faced with a similar problem of matching 2D data to a 1D model, but with endpoints only; we therefore discuss parallels between our work and existent work in Fitts' law literature when we found them. [4]

4.1 Parametrization of movement

We introduce the following quantities to parametrize any trajectory (see Fig. 3):

- The start location $O = (x_0, y_0)$, defined by the cursor location at time $t = 0$;

[4] A naïve solution would consist in using the remaining euclidean distance to the target as proxy signal upon which to compute PVPs. However, this would necessarily produce an underestimate of the spread of data, since undershoots and overshoots would in that case not be distinguished, as recognized previously by Wobbrock *et al.* [39] in the case of endpoints. A simple workaround to differentiate undershoots from overshoots is to consider the distance from the starting point rather than the distance to the target. Although this solves the previous issue when the movement is always perfectly aligned with the start-target axis (like in the 1D case), it does not in all other cases.

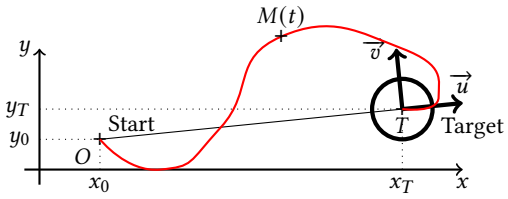

Fig. 3. Origin of movement (0), target (T), and current position ($M(t)$) of the cursor. The orthogonal basis used to define α and β is also given.

- The target location $T = (x_T, y_T)$, which remains fixed throughout the movement;
- The primary direction of movement defined as $\overrightarrow{OT}$ is normalized with the euclidean norm to $\overrightarrow{u} = \overrightarrow{OT}/||\overrightarrow{OT}||_2$. This vector is dependent on the initial conditions only.
- The direction orthogonal to movement is defined such that $\{\overrightarrow{u}, \overrightarrow{v}\}$ form the standard orthonormal basis in the plane.

We define $\overrightarrow{MT}$ as the vector relating the current position $M(t)$ to the (fixed) target position T at each time, and compute its projections $\alpha(t)$ and $\beta(t)$ on the basis $\{\overrightarrow{u}, \overrightarrow{v}\}$

$$\alpha(t) = \overrightarrow{MT} \cdot \overrightarrow{u}, \tag{11}$$

$$\beta(t) = \overrightarrow{MT} \cdot \overrightarrow{v}. \tag{12}$$

In the 1D case, $\beta = 0$ and $x(t) = \alpha(t)$.

A set of trajectories from pilot data in the $\{\overrightarrow{u}, \overrightarrow{v}\}$ plane is illustrated Fig. 8 (Appendix). It shows that the $\beta(t)$ component of movement is roughly symmetric. This suggests that the experimental protocol used to acquire this pointing data (presented in the next section) successfully balanced the initial directions of movement.

4.2 Accounting for all directions of movement

To generalize the notion of variance to the multidimensional setting, we use the covariance matrix, which gives the joint variability between each pair of elements in the (α, β) vector. The vector $\overrightarrow{MT}(t)$ has covariance matrix $\Gamma(t)$

$$\Gamma(t) = \begin{bmatrix} \Gamma_{\alpha\alpha}(t) & \Gamma_{\alpha\beta}(t) \\ \Gamma_{\alpha\beta}(t) & \Gamma_{\beta\beta}(t) \end{bmatrix}. \tag{13}$$

$\Gamma_{\alpha\alpha}(t)$ is the variability of the α component with itself, *i.e.*, it is the (univariate) variance of the α component. We can estimate each term of the covariance matrix with the sample covariance estimator

$$\Gamma_{\alpha\beta} = \frac{1}{N-1} \sum_{i=1}^{N} (\alpha - \mu_\alpha)(\beta - \mu_\beta), \tag{14}$$

where μ_α (resp. μ_β) is the *sample* average for α (resp. β) and where i indexes trajectories. For $\Gamma_{\alpha\alpha}$, we recognize the classical (univariate) sample variance. We have also dropped the time variable (t) for lighter notations here, and so will we for all similar quantities in the remainder of the paper.

4.3 Candidate 2D PVPs

The *variance model* [16] predicts the evolution of the (univariate) standard deviation, but the data in the 2D setting is multivariate. Our strategy here is to apply the one-dimensional *variance model* to a one-dimensional representation $\tilde{\sigma}^2$ of the multivariate covariance matrix. In line with the existing statistical literature on multivariate analysis [33], we take $\tilde{\sigma}^2$ to be either

$\tilde{\sigma}^2_\alpha$**:** variance only in the direction of movement, which gives PVP_α;[5]
$\tilde{\sigma}^2_{\text{total}}$**:** the total variation, which gives $\text{PVP}_{\text{total}}$;
$\tilde{\sigma}^2_{\text{generalized}}$**:** the generalized variance, which gives $\text{PVP}_{\text{generalized}}$.

We propose the three following PVPs, where $\tilde{\sigma}$ denotes the time series of the profile:

PVP_α: Projection in the direction of movement. Here, we consider the standard deviation only in the direction of movement[6]

$$\tilde{\sigma}_\alpha(t) = \sqrt{\Gamma_{\alpha\alpha}}. \tag{15}$$

The same calculation for spread is performed for Fitts' law experiments —who deal with endpoints only— in the standard ISO procedure [37].

Note that the ISO standard provides no justification for this method, and even though it is known that errors orthogonal to the principal direction of motion are usually of smaller magnitude than in the direction of motion [26], it is unclear why deviations in the orthogonal direction should be ignored altogether [39]. One possible explanation could be the *minimum intervention principle* [38], which states that deviations that do not interfere with the task performance are not corrected by the human sensorimotor system. The next PVPs do account for deviations in all directions.

$\text{PVP}_{\text{total}}$: trace of the covariance matrix. Here, we consider the sum of the variances in the direction of movement and orthogonal to the direction of movement, *i.e.*, the sum of the diagonal elements of the covariance matrix, also called the *trace* Tr of the covariance matrix

$$\tilde{\sigma}_{\text{total}}(t) = \sqrt{\text{Tr}(\Gamma)} = \sqrt{\Gamma_{\alpha\alpha} + \Gamma_{\beta\beta}}. \tag{16}$$

This quantity is known as the total variance (or variation), and it measures the average euclidean distance of the trajectories to the mean trajectory [8]. One of its advantages is that it accounts for variance in the direction of movement as well as its orthogonal direction. When plugging in the sample estimator Eq. 14, we recognize that

$$\tilde{\sigma}_{\text{total}}(t) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[(\alpha-\mu_\alpha)^2 + (\beta-\mu_\beta)^2\right]}, \tag{17}$$

which is what Wobbrock *et al.* [39] coined the 2D distance-from-centroid formula in the case of endpoint deviation.[7]

$\text{PVP}_{\text{generalized}}$: Determinant of the covariance matrix. The determinant of a matrix Γ, $\det(\Gamma)$, is a function of the matrix's entries which occurs throughout mathematics. One of its interpretation is as the surface area generated by the matrix; its square root is then the characteristic amplitude of the spread of the data. We can actually go beyond the

[5] If $\Gamma_{\alpha,\beta} = 0$, then this is also known as the spectral norm of the covariance matrix *i.e.*, it is its largest eigenvalue. In practice, we have $\Gamma_{\alpha,\beta} \ll \Gamma_{\alpha,\alpha}$ and so PVP_α is almost equal to the spectral norm.
[6] We similarly define PVP_β, by using β instead of α everywhere.
[7] The normalized version $\sqrt{\Gamma_{\alpha\alpha}/2 + \Gamma_{\beta\beta}/2}$ might have slightly better properties [33], but we stick with the unnormalized version to match Wobbrock's distance-from-centroid formula.

simple intuition: the n-th root of the determinant of an n-dimensional matrix is known as Wilk's generalized variance and can be shown to have properties desirable for spread measures [8, 33]. We thus take as our final univariate measure of spread[8]

$$\tilde{\sigma}_{\text{generalized}}(t) = (\det(\Gamma))^{1/4} = \left(\Gamma_{\alpha\alpha}\Gamma_{\beta\beta} - \Gamma_{\alpha\beta}^2\right)^{1/4}, \tag{18}$$

This formulation is the only one of the three to account for cross-correlation terms ($\Gamma_{\alpha\beta}$) and has to our knowledge no equivalent in the endpoint-only case. We also further show in the next Subsection that $\text{PVP}_{\text{generalized}}$ can be interpreted as the average of PVP_α and PVP_β.

N-dimensional PVPs. The definitions for PVPs that we introduced generalize to any dimension N. The procedure is as follows: first construct a direct orthonormal basis whose first basis vector is aligned with the direction of movement, on which each trajectory should be projected. Then compute the covariance matrix Γ for each point in time and either 1) compute the square root of the first component of Γ for PVP_α, or 2) compute the square root of the trace of Γ for $\text{PVP}_{\text{total}}$, or 3) compute the $2N$-th root of the determinant of Γ for $\text{PVP}_{\text{generalized}}$.

Other PVPs. We have selected and rationalized three PVP computations. There are opportunities for others, say, $\sqrt{\Gamma_{\alpha\beta}}$ or $\sqrt{\Gamma_{\alpha\alpha} + \Gamma_{\alpha\beta} + \Gamma_{\beta\beta}}$, but at this point, we are not interested in systematically investigating the "best" 2D-PVP.[9]

4.4 Theoretical Comparison of 2D PVPs

Before an empirical study, we can compare PVPs on formal grounds. First, it is obvious from Eqs. (15) and (16) that $\text{PVP}_{\text{total}}$ will always be larger than PVP_α.

Then, we found by observing the data presented in the next section, that the cross-covariance term $\Gamma_{\alpha\beta}$ is usually quite small compared to the variance terms. Making the approximation that $\Gamma_{\alpha\beta} \sim 0$, the determinant of the covariance matrix thus reduces to $\Gamma_{\alpha\alpha}\Gamma_{\beta\beta}$. When taking the logarithm of the quartic root and using the properties of the logarithm, one gets:

$$\log(\Gamma_{\alpha\alpha}\Gamma_{\beta\beta})^{1/4} = \frac{1}{2}\left[\log\sqrt{\Gamma_{\alpha\alpha}} + \log\sqrt{\Gamma_{\beta\beta}}\right]. \tag{19}$$

Hence, $\text{PVP}_{\text{generalized}}$ is obtained by averaging PVP_α and PVP_β in the log-lin plane. The well-known inequality of arithmetic and geometric means also implies that $\text{PVP}_{\text{generalized}}$ is always smaller than $\text{PVP}_{\text{total}}$.[10]

If we further assume that $\Gamma_{\beta\beta} \leq \Gamma_{\alpha\alpha}$, which is reasonable everywhere except at the very end of the movement, since variability in the direction orthogonal to movement is usually lower than in the direction of movement, we have

$$\text{PVP}_{\text{generalized}} \leq \text{PVP}_\alpha \leq \text{PVP}_{\text{total}}. \tag{20}$$

The magnitude of these differences is due to the difference in magnitude between $\Gamma_{\alpha\alpha}$ and $\Gamma_{\beta\beta}$, which, we believe, is in part related to the choice of experimental protocol, especially whether the direction of approaches are balanced or not. The inequality is illustrated in Fig. 9 (Appendix) on real data.

[8] For a more precise definition, we refer to any linear algebra textbook.

[9] And as the results will show, there may be practically little interest in selecting either one of them in a "typical" situation. Differentiating between the profiles would likely be interesting in situations where variance is large and/or unusually high in the orthogonal direction.

[10] The inequality reads $\sqrt{xy} \leq \frac{x+y}{2}$, which gives $\frac{1}{2}\log\sqrt{\Gamma_{\alpha\alpha}\Gamma_{\beta\beta}} \leq \frac{1}{2}\log\frac{\Gamma_{\alpha\alpha}+\Gamma_{\beta\beta}}{2}$. The left hand side is $\text{PVP}_{\text{generalized}}$, while the right hand side is always smaller than $\text{PVP}_{\text{total}}$.

4.5 PVPlib: a Python library

We released a Python library called PVPlib to make PVPs accessible to the community. The library features 3 classes that correspond to the 3 PVPs mentioned before. These classes act as containers, to which trajectories can be added, and from which PVPs can be computed, and fitted based on the procedure described in Sect. 3. These classes are N-dimensional, and automatically infer the dimensionality of the trajectories. The library performs some preprocessing on the trajectories before computing PVPs, which includes resampling, filtering and resynchronization of trajectories, and comes with plotting features. The library can be downloaded directly from PyPI https://pypi.org/project/pvplib/ and has a GitHub public repository which also hosts its documentation at https://github.com/jgori-ouistiti/PVPlib.

5 EMPIRICAL STUDY: METHOD

We report on a controlled experiment to validate 2D-PVPs, where we investigate whether:

(1) Results from the 2D-PVP match those of the 1D-PVP in Gori and Rioul's work [16]. In particular, we hypothesize that
   - The 2D-PVP profiles display three phases: A first phase where the PVP increases rapidly, followed by a second phase where the PVP decreases at an exponential rate, and a third phase where the PVP is constant.
   - Distance significantly affects τ, D_τ, and σ_0, but not C, and with considerable effect sizes.

(2) Comparisons performed with a traditional Fitts' law experiment are equivalent to comparisons performed with a new protocol analyzed with 2D PVPs. We namely investigate whether, in a classical protocol, linear regression parameters can be predicted with features of the 2D-PVPs.

We performed a single controlled experiment to acquire data for this investigation, as described below.

5.1 New Protocol

We devised a new dual-minimization protocol, where the participant aims towards a very small dot which appears at a given distance and a pseudo-random angle. We selected the smallest dot possible that was still attainable with reasonable effort (less than 1.33 mm wide). Participants were instructed to select the target by emphasizing accuracy, but taking no more time than needed. We used a pseudo-random method to place targets, to remove prediction[11] and learning effects. The protocol we propose is essentially a balanced, reciprocal, 2D version of Guiard's dual-minimization protocol [24, 25].

5.2 Procedure

We conducted a two factor full factorial design, with factors Protocol × Device, where Protocol is either *ISO* or *DualMin* and Device is one of {*Mouse, Touchpad, Controller*}. In the *ISO* condition, participants were instructed to perform the ISO ring of rings task as quickly as possible while aiming for an error rate of about 4%, as prescribed by the ISO standard [37]. In the *DualMin* condition, participants were instructed to perform the new protocol.

Each participant performed 6 blocks, where each block is a combination of Protocol × Device. To counterbalance order and carry-over effects for our 12 participants, we used two (different) 6-by-6 Latin squares. To keep the experiment length less than 30 minutes per participant, we did not use a full factorial design for D and W in the *ISO* condition. Instead, we determined a set of 6 (D,W) pairs that were randomly shuffled for each new block: $\{(200, 124), (440, 66), (1160, 90), (920, 38), (680, 18), (1400, 14)\}$ (in pixels), resulting in ID's $\{1.4, 2.9, 3.8, 4.7, 5.3, 6.7\}$. These pairs were determined to

[11] This happens if the participant already knows where the next target is going to appear.

satisfy a multi-objective optimization problem with following aims: a) low correlation between D and W, to estimate effects and associated standard errors of D and W properly [15] b1) equally spaced ID, and b2) maximum range of ID spanned, to have a good range for the covariates c) reasonable values for D and W (no risk of hitting the edge of the screen, target attainable). More details can be found in the Supplementary material. The 6 D conditions were similarly shuffled in the *DualMin* condition.

The experiment starts with a training session that ends when the participant signals that they are accustomed to the setup. For each condition, and for each (D,W) pair, participants where asked to perform 21 trials. The first trial was systematically dropped. Optional 1-minute breaks where offered to participants between each condition (each condition consisting of 126 movements —which take between 1 and 2 second per movement for most participants— this amounts to a potential 1-minute break every 2 to 5 minutes).

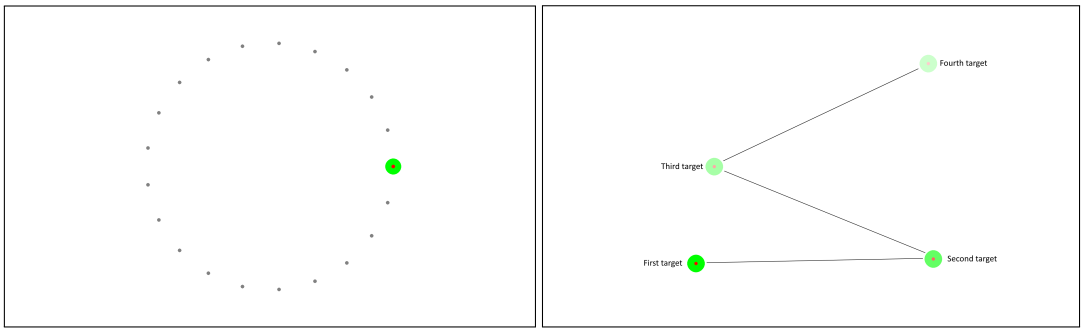

Fig. 4. Left: Task in the *ISO* condition. All targets are visible. We have represented the condition with the smallest W. Right: Task in the *DualMin* condition. Only one target is visible at a time during the experiment, but 4 images have been stacked here. For both panels, the user should aim towards the red target; the green disk is only here to facilitate finding small targets which may be hard in the *DualMin* condition.

5.3 Apparatus

Participants were seated on a chair during the experiment, which they could adjust as needed. The experiment was performed on a Windows Microsoft Surface Pro 4 (2736x1824), running PopOS with the Linux Surface Kernel. Participants were asked to use a standard commercial external mouse, the Surface's native touchpad or an external Xbox controller in a random order. In the case of the mouse and touchpad, the native cursor acceleration was enabled, and sensitivity was set to PopOS's default setting. For the controller, we used a custom transfer function, as we judged the default on PopOS to be poor in a pilot study.

In the *ISO* condition, participants were shown a gray arrangement of circles, one of which was red to indicate the target to select. In the *DualMin* condition, participants were shown only a red target. In both cases, a larger green circle surrounded the target, so that participants wouldn't spend time searching for the target, see Fig. 4.

5.4 Data Collection

Our software captured cursor position at 100Hz, and registered full trajectory information (position and time), target information, and bookkeeping information. In total, we logged 9072 movements: 12 PARTICIPANTS × 6 (D,W) PAIRS × 6 (PROTOCOL × DEVICE) × 21 TRIALS. This dataset is available online at https://github.com/jgori-ouistiti/data_chi23. All distances are expressed in mm.

5.5 Data Analysis

5.5.1 2D-PVP profiles. We computed PVP profiles for each block using PVPlib (see Subs. 4.5). First, we discarded the first trajectory of each block. To synchronize the trajectories, we started by re-interpolating the time series at 100 Hz to account for unequal sampling times. We then applied a low-pass filter on the time series (kaiser window) with cut-off frequency at 10 Hz. Trajectories were then pooled (per condition), and extended. Finally, we computed PVP_α, $\text{PVP}_{\text{total}}$, and $\text{PVP}_{\text{generalized}}$ as described in Section 4.

5.5.2 Effect of task parameters on 2D-PVP features. To investigate the level of association between PVP features and task parameters, we ran linear multilevel regressions [15] on τ, D_τ, σ_0 and C using *lme4* [2]. These models allow fitting an intercept and slope per device while controlling for the variance induced by multiple participants and the dependency of data in repeated measurements. We defined D, W and the interaction $D \times W$ as fixed effect. We also controlled for dependence of data by adding a random effect Participant and controlled the variation due to Device when needed. The various statistical models we fitted are described using the syntax of R formulae. An example formula is `tau ~ D*W + Device + Device:D + (1|Participant)`. Here, `Device` is a categorical factor with n categories, and will add $(n-1)$ constants in the model (the baseline device being integrated in the model's intercept), `Device:D` is an interaction term which will add $(n-1)$ device-dependent slopes (the baseline device being in the model's slope), `(1|Participant)` adds a random intercept per participant, and `D*W` is equivalent to `D + W + D:W`.

5.5.3 Predicting Fitts' law parameters. We estimated nominal and effective version of Fitts' law parameters for each device separately. We then estimated the intercept and slopes for each device based on our predictive formulae Eqs. (9) and (10).

5.6 Participants

We recruited 12 participants, based on a power analysis, where we used data from the Pointing Dynamics dataset [32] as pilot data. We extracted PVPs and corresponding parameters with our library, on which we fit a general linear multilevel model as described above. For our target effect size, we used a conservative estimate, namely the lower bound of the 95% confidence interval of the effect size of the coefficient associated with the D covariate for regressions on τ, D_τ, σ_0. That means that even if D's effect turns out to be 2 standard deviations smaller than what was estimated from the pilot data, we would still have the appropriate statistical power to detect it. To computer power curves, we used `simr` [20]. Taking the least favorable of the three regressions, with a target power $\beta = 0.8$, yields $N = 9$ participants as our target sample size. We ended up recruiting 12 participants to make balancing the experiment design, which has $2 \times 3 = 6$ levels, easier. The full analysis can be found in the supplementary material.

6 EMPIRICAL STUDY: RESULTS

In total, we computed 3 PVP × 432 blocks (12 × 6 × 6) = 1296 - 4 = 1292 profiles (4 missing due to a bug in the software).

6.1 Data presentation

PVP_α for the 6 Device × Protocol conditions for a single participant are displayed Fig. 5. The mouse produces the cleanest profiles, with PVPs quite packed, the controller produces very long and gentle but noisy slopes in the second phase, while the touchpad has profiles that are affected by D much more than the others, as visible by how PVPs are separated. We also see the difference in target sizes in the *ISO* condition, where the PVPs level off at different values of

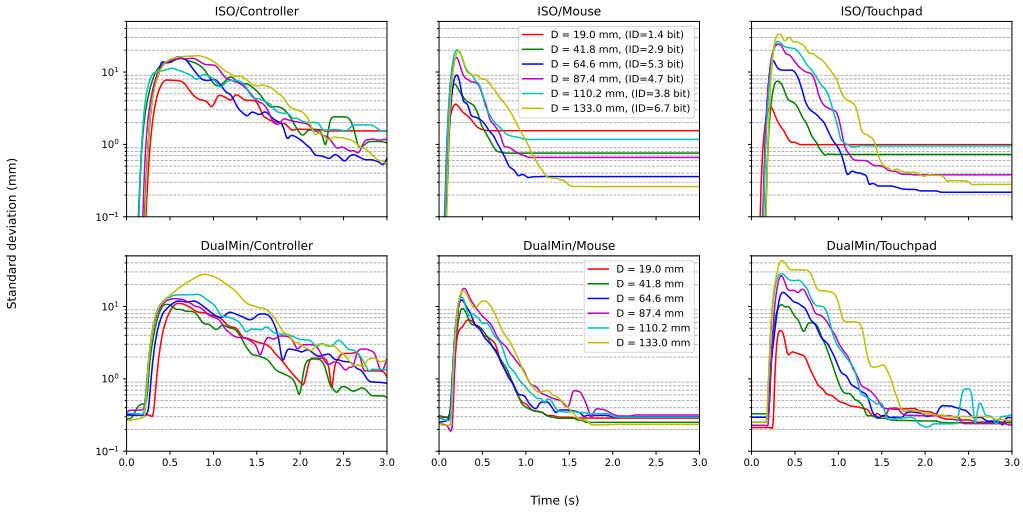

Fig. 5. PVP_α for Participant 136418 for the 6 Protocol × Device conditions. Each plot displays the PVP (standard deviation in mm) against time (s) with identical scales.

standard deviation. Overall, the profiles are more noisy than what was presented in our prior work [16]; this is likely due to the fact that we only have 20 movements per block here, instead of about 40 to 50 movements per block. Also, the controller and touchpad produce more variable trajectories than the mouse; this is especially true for the controller.

6.2 Three phases

To identify the three phases of the profiles, we fitted each profile as described in Subsection 3.3. We determined that 8 out of 1292 profiles were not a good match with the three-phase description (either $r^2 < 0.8$, or those who produced a positive slope (C < 0)). The bad profiles where all in the *ISO* protocol, mostly for the smallest ID (5/8 for ID = 1.4).

6.3 Comparison between PVPs

A visual comparison of the three PVPs is shown Fig. 9 (Appendix), which illustrates Eq. (20): PVP_{total} is always above the others, while PVP_α is usually above $PVP_{generalized}$. Overall, PVP_α follows PVP_{total} during the early parts of the profiles and then crosses over to $PVP_{generalized}$. In the remainder of the work, we present results only for PVP_α: effect sizes may be marginally different between PVPs, but no significant difference emerged between the profiles in our analyses. After visualizing the data, we fit a linear multilevel model to each of τ, D_τ, σ_0 and C, with the goal of estimating effect sizes. We won't be focusing on statistical significance: given the large sample size (N=428 = 432-4 after removal of bad trials) many of the effects we estimate will be significant.

6.4 Effects of D on τ, D_τ, σ_0 and C

We first visualize the effects of D on τ, D_τ, σ_0 and C via box plots grouped by device in Fig. 6, since we hypothesize that different devices may lead to different phase features (box plots for the effects of W and ID are shown in the

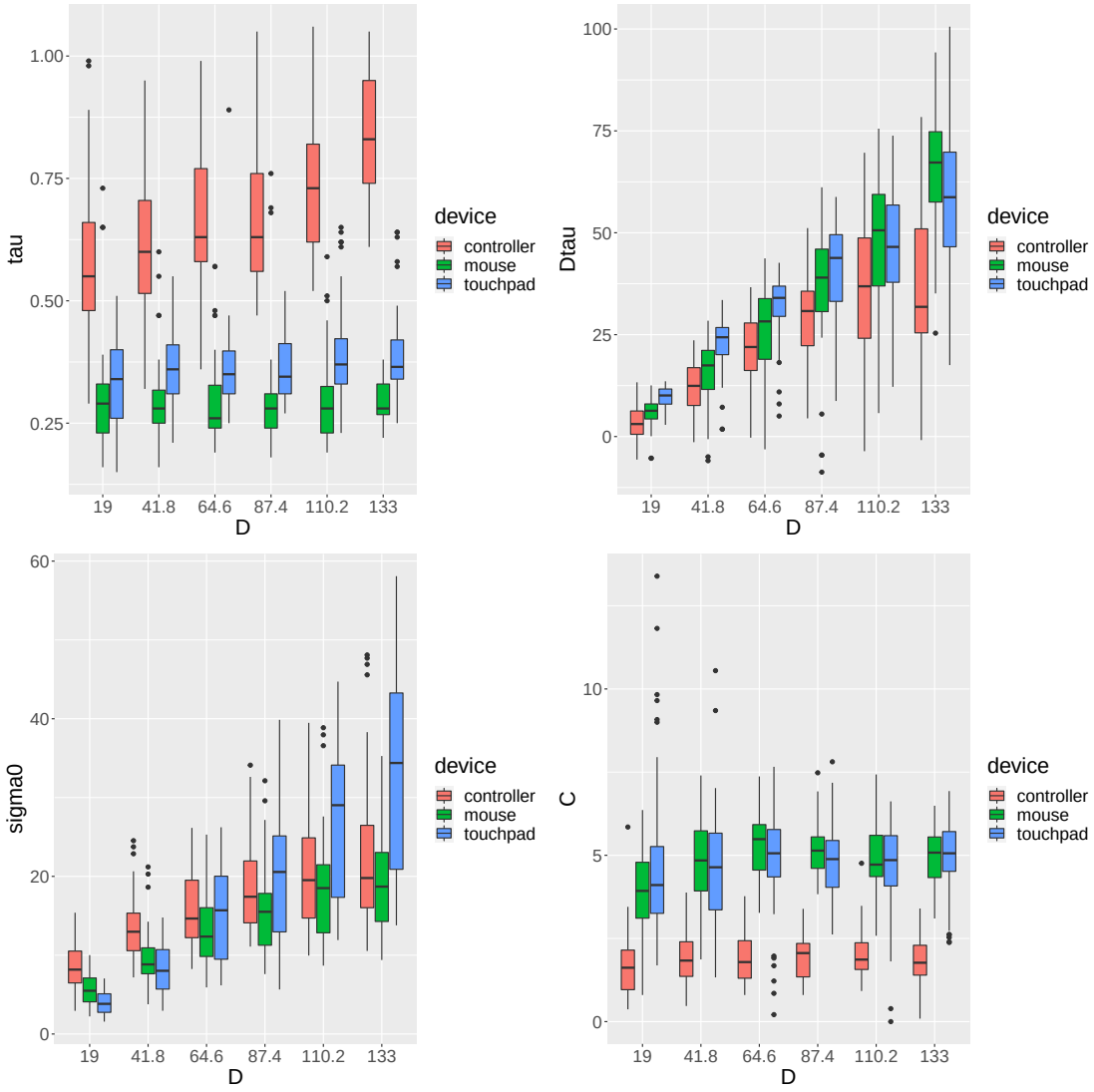

Fig. 6. Box plots grouped by device, showing the effect of D on: top left – τ, top right – D_τ, bottom left – σ_0, bottom right – C. Note that D is given in mm for comparison with other studies.

Supplementary material: they show much smaller effect sizes, as confirmed by Tab. 1, even though some effects may be statistically significant, see Tab. 3).

Results show that

- τ is about constant for the mouse and touchpad, with the touchpad leading to somewhat larger τ. The effect of D on τ is however marked for the controller, for which τ is also on average much larger.
- D_τ shows a strong increase with D. We see that the increase levels off for the controller, which suggests the slope for the controller will be different from the other devices.

- For σ_0, the case is similar to D_τ, although this time slope for touchpad is quite different from the two other devices.
- C seems constant across values of D. There is a notable difference between the controller, which has values for C two to three times lower than the mouse and touchpad.

We then used multilevel models to estimate the effect sizes of D (see Subs. 5.5.2). Inclusion of random effects is based on the box plots see Fig. 6 and the Supplementary materials. The estimated effect sizes are given Tab. 1, and the full fit results are delayed to Tab. 3 in the Appendix.

Table 1. Estimated effect sizes for PVP features. For each covariate, we provide the estimated effect size, as well as average contribution to τ in percentage.

	Mean	**Intercept**	**D**	**W**	**mouse**	**touchpad**	**D × W**	**D × mouse**	**D × touchpad**
$\boldsymbol{\tau}$ *(s)*	0.470	0.577 (123%)	0.003 (49%)	-0.013 (-9%)	-0.222 (-47%)	-0.191 (-41%)	$3.5e^{-5}$ (2%)	-0.003 (-49%)	-0.003 (-49%)
$\boldsymbol{D_\tau}$ *(mm)*	32	2.13 (7%)	0.246 (60%)	-0.330 (-4%)			0.003 (2%)	0.232 (56%)	0.218 (52%)
$\boldsymbol{\sigma_0}$ *(mm)*	17.7	4.30 (24%)	0.16 (69%)	-0.013 (-3%)			$2.1e^{-4}$ (0%)	-0.023 (-10%)	0.089 (38%)
$\boldsymbol{C}$ *(bit/s)*	4.11	1.65 (40%)	0.004 (7%)	0.064 (1%)	3.06 (75%)	3.25 (79%)	-0.001 (-6%)		

6.4.1 *Duration of the first phase: τ.* We fitted the linear model

```
tau ~ D*W + (1|Participant) + device + device:D,
```

where we estimate the effects of D, W, and their interaction `D*W` on τ. We controlled for participant differences by adding a random intercept `(1|Participant)`, and allowed different intercepts (`device`) and slopes (`device:D`) for each device. This modeling choice was motivated by the controller (see the top left panel of Fig. 6), which shows a higher average value and slope.

The intercept contributes by far the most to the value of τ, ($\tau_{\text{controller}} = 0.58\,\text{s}$, $\tau_{\text{mouse}} = 0.36\,\text{s}$, $\tau_{\text{touchpad}} = 0.40\,\text{s}$), with other effects being small, except for the slope in the controller[12].

6.4.2 *Distance covered during the first phase: D_τ.* We fitted the linear model

```
Dtau ~ D*W + (1|Participant) + device:D.
```

Here, we don't introduce a different intercept per device, because the average values look comparable in the top right panel of Fig. 6. We do introduce a different slope per device because one can see that the slope of the controller seems to reach a threshold[13] There is a strong effect of D on D_τ, which varies per device: the effect of D is about twice as strong for *mouse* and *touchpad* compared with *controller*, while other effects contribute marginally ($D_{\tau,\text{mouse}} = 0.48\text{D}$, $D_{\tau,\text{touchpad}} = 0.47\text{D}$, $D_{\tau,\text{controller}} = 0.25\text{D}$).

6.4.3 *Standard deviation at the end of the first phase: σ_0.* We fitted the linear model

```
Sigma0 ~ D*W + (1|Participant) + device:D,
```

[12] D × *mouse* represents the difference in D slope between the baseline group (*controller*) and the *mouse* group. Summing the two slopes gives a null slope, showing that D has virtually no effect on τ in the *mouse* group. The same is true for the *touchpad* group

[13] This threshold in itself is quite interesting, and seems to suggest that the controller's transfer function has too small of a gain for high speeds.

as the slope of the touchpad seems much larger (see bottom left panel of Fig. 6). Average values again look rather similar, so we did not introduce a different intercept for each device. There is a strong effect of D on σ_0, with D having a stronger effect for *touchpad* than the other two devices ($\sigma_{0,\text{controller}} = 4.3+0.16\text{D}$, $\sigma_{0,\text{mouse}} = 4.3+0.14\text{D}$, $\sigma_{0,\text{touchpad}} = 4.3+0.25\text{D}$). The intercept contributes significantly, accounting for about 30% of σ_0's average value.

6.4.4 Participant performance: C. We fitted the linear model

```
C ~ D*W + (1|Participant) + device.
```

Because the slope of C looks to be null, but there are obvious average difference between devices (see bottom right panel of Fig. 6), we only fitted a different intercept per device. The only strong effects are the intercept's, where C for the controller is much lower than the other two devices ($\text{C}_{\text{mouse}} = 4.72$ bit/s, $\text{C}_{\text{touchpad}} = 4.87$ bit/s, $\text{C}_{\text{controller}} = 1.66$ bit/s).

6.5 Comparison with Fitts' law

As explained in Sect. 3 both the slope and intercept of Fitts' law can be predicted from the PVPs' features. To verify this prediction, we first fitted Fitts' model, with separate intercept and slopes for each device and for both the nominal index ID (Eq. (1)) and the effective index ID_e (Eq. (2)).

```
MT ~ IDe + device + device:IDe
MT ~ ID + device + device:ID
```

The parameters estimated via the linear model are shown Fig. 7 and Tab. 2 (also see Tab. 4 in the Appendix).

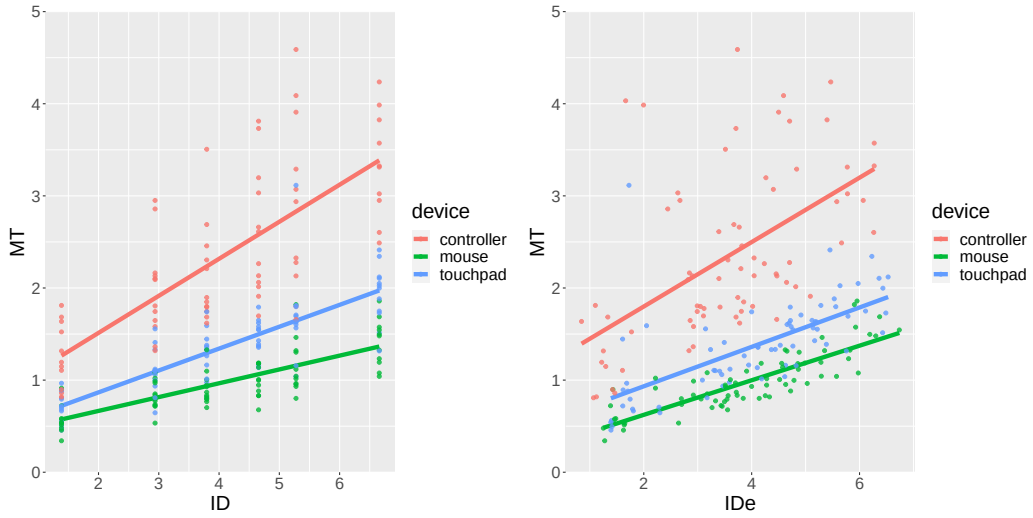

Fig. 7. Fitts' law in nominal (left) and effective (right) versions fitted on data from the *ISO* condition.

We then estimated the intercept and slope $\hat{a}$ and $\hat{b}$ for the nominal and effective laws based on Eqs. (9) and (10), and effect sizes from Tab. 1. The predictions are accurate for the mouse and touchpad both in the effective and nominal versions of the law for the intercepts (with a maximum error of about 120 ms for the touchpad in the nominal version). For the slope estimates, the estimates are particularly precise for the effective version of the law (error of 0.02 s/bit for

Table 2. Estimated parameters for nominal and effective Fitts' law using the classical method, compared with $\hat{a}$'s and $\hat{b}$ estimated from Eqs. (9) and (10).

	ε	**Nominal b**	**Effective b**	$\hat{b}$	**a Effective**	$\hat{a}$ **Effective**	**a Nominal**	$\hat{a}$ **Nominal**
mouse	0.8%	0.15	0.19	0.21	0.25	0.21	0.37	0.27
controller	2%	0.40	0.35	0.54	1.10	0.60	0.71	0.58
touchpad	0.8%	0.24	0.21	0.21	0.51	0.41	0.39	0.51

the mouse, below 0.005 s/bit for the touchpad), but a little less precise for the nominal version (0.04 s/bit error for the mouse, 0.03 s/bit for the touchpad). The predictions are less accurate for the controller (about 0.2 s/bit for the slope, and half a second of error for the intercept in the effective version). The fact that results found using PVPs can be compared to those with Fitts' law is important considering the prevalence of Fitts' law in the prior literature, and may help researchers using PVPs benchmark prior work.

7 DISCUSSION

7.1 Two(three) phase model

The controlled experiment supported that the two[14] phase model was applicable in almost all blocks. We included two popular devices (mouse and touchpad), as well as a less popular one, with a custom transfer function, to evaluate this property broadly. The fact that even with the controller we clearly see the two phases supports the idea that this property is due to a low-level human process, and suggests it will be applicable in many HCI situations.

Two-phase models are actually widespread in the study of human movement, starting with Woodworth [40], who hypothesized that a ballistic phase primarily used to cover distance was followed by a second homing-in phase. This two-phase description was further developed, particularly by Elliott and colleagues [9–11] who suggested that the two phases were a feedforward control phase followed by an online (feedback) control phase.

PVPs are aligned with that description: the first phase of the PVP is a fast distance covering mechanism in which variance increases — properties usually associated with feedforward control [3, 14] — while the second phase of the PVP is a longer one based on a variance reduction model with feedback information.

7.2 Which PVP version should we use?

Our study showed little difference between the three PVP versions that we proposed. Most of the variance is along the $\Gamma_{\alpha\alpha}$ component at the start of the movement, gradually becoming closer to the $\Gamma_{\beta\beta}$ component, as shown by the fact PVP_α follows $\text{PVP}_{\text{total}}$ at the start of the movement, but then ends up following $\text{PVP}_{\text{generalized}}$ towards the end of the movement. While we could have selected a criterion to compute a score and select the "best one", we feel there is little sense in doing that: we believe that the user of PVPs should reflect which variance component they want to account for ($\Gamma_{\alpha\alpha}, \Gamma_{\beta\beta}, \Gamma_{\alpha\beta}$) in order to determine which one is the most appropriate. We suppose that differences in PVP versions will be meaningful only for data with high variance, in particular when in the direction orthogonal to movement.

[14]We intentionally play on the fact that while the model is a two-phase one, the method introduces a third phase for the purpose of estimation.

7.3 Effect of task parameters on PVP features

Overall, we replicated the results from the 1D study [16]. All three devices shared the same predominant effects of D on PVP features. The only discrepancy that we found with [16] is an intercept on σ_0. These features inform us on the device in more detail than what can be achieved with only the intercept a and slope b from a traditional Fitts' law experiment. For example, in comparison with the mouse and the touchpad, for the controller, τ was much higher, while D_τ was lower: the first phase, while being longer, covered less distance. This suggests that the transfer function of the controller could be better tuned (we believe that the maximum speed of the cursor was too low, and hence a higher gain in the high speed condition would be beneficial).

7.4 Link with Fitts' law experiments and parameters

At a high level, our results paint the same picture between Fitts' law experiments and PVPs: performance between the touchpad and the mouse are similar, while the touchpad scores comparatively lower. Interestingly though, the protocol that we used is much simpler because it eliminates two factors: instead of crossing D and W as in Fitts' law experiments, we only need a condition with a very small target and large distance for the *DualMin* protocol. In this work, we used 6 different conditions in the *DualMin* protocol, but this was only to provide enough values of D to validate the properties of Section 3 for 2D PVPs: a single condition would have been enough. We were also able to predict with a good accuracy the parameters of Fitts' law parameters from the PVP features, despite the predictive formulae incorporating highly non-linear terms (*e.g.*, $\log_2(\mathrm{erf}^{-1}(1-x))$). To our knowledge, this is the first time a model is able to predict the values, and not just the shape associated with Fitts' law.

8 IMPLICATIONS FOR FITTS' LAW AND FITTS' LAW EXPERIMENTS

In the introduction, we listed some open problems and discussions regarding Fitts' law experiments. Below, we revisit that list under the light of our results.

8.1 Which version of ID should we use, and why?

Eqs. (9) and (10) suggest the nominal index $ID = \log_2 \frac{D}{W}$ and the effective index $ID_e = \log_2 \frac{D}{\sigma_\infty}$ where σ_∞ is equivalent to σ in Eq. (2). Thus, the Variance Model [16] is consistent with the received expression for ID except for the "+1" term inside the logarithm. How much difference does that "+1" cause? As previously discussed [16, 27, 41], that term is of little interest, because it induces changes only for very low values of the ratio D/W where Fitts' law is known to be a poor model [6]. In fact, for low D/W, the proper approach would be to better model the first phase, and build a model based on that. As a result, we suggest using either version for ID > 2 or 3, but suggest further work is needed for the small ID regime.

The *Variance Model*, via Property 6 also suggests a *local* index of difficulty id, valid between any two instants t_1 and t_2 in the variance decreasing phase

$$\mathrm{id} = \log_2 \frac{\sigma(t_1)}{\sigma(t_2)}, \tag{21}$$

According to Guiard [23], ID runs on a non-ratio scale of measurement. In contrast, the local id is a well-defined ratio between two measurable standard deviations.

8.2 What is the interpretation of ID?

ID or id are, contrary to the received interpretation [37], not expressed as channel capacities; instead they measure the information that needs to be transmitted to *reliably* perform the aiming task, independently of the distance covering mechanism. So, rather than being "rates", ID and id are "loads", which actually matches the notion of difficulty inherent to the name "index of difficulty": a more difficult task requires more information to be transmitted. This interpretation shows that MacKenzie's argument for adding "+1" to the logarithm "so that $\log_2\left(1+\frac{D}{W}\right)$ better matches the formula for the Shannon capacity" [29] is thus flawed, since ID is not a capacity in the first place.

8.3 How should the intercept be interpreted?

The intercept of Fitts' law can be expressed from the PVP features, see Eqs. (4), (6), and [16, Eqs. (27), (29)]

$$a = \tau_0 + \alpha \mathrm{D} + \frac{1}{C}\log_2(2\sqrt{2}\gamma \times \mathrm{erf}^{-1}(1-\varepsilon)) \qquad \text{(nominal)}, \tag{22}$$

$$a = \tau_0 + \alpha \mathrm{D} + \frac{1}{C}\log_2(4.133\gamma) \qquad \text{(effective)}, \tag{23}$$

where c and d are some functions. Thus, τ is the contribution of

(1) An incompressible positive time τ_0, which differs from one participant to another (see Eq. (4)).
(2) A scaling effect due to distance, via the term α in Eq. (11), which depends on the device used (see Device:D in Tab. 1).
(3) A second scaling effect, due to γ, that relates the standard deviation at the end of the first phase to D. This effect also includes:
   - the arbitrary 4.133 constant in case of the effective law;
   - the miss rate via $\mathrm{erf}^{-1}(1-\varepsilon)$ in case of the nominal law.

Judging from the many contributions to the intercept, it becomes clear why interpreting the intercept is so hard [41] even if excluding time due to extra cognitive processes. However, contrary to what is typically argued [37, 41], there seems to be no reason to expect a positive or null intercept: in fact negative intercept may be predicted, for example if the effect of γ is very small.

This expression for intercept is also interesting, in that it expresses what Guiard [22] calls the *weak version* of Fitts' law, the strong version being Eq. (1) with constant intercept. In the weak version, an effect of scale shows up, and MT has a term linear in D. For some experiments, such as Fitts' disc-transfer experiment [12], that effect is noticeable. The expression for the intercept shows that both the weak and strong versions of Fitts' law can be predicted (depending on the value of α).

8.4 Can throughput be unequivocally defined?

The throughput is an effective rate of information transmission; given a high effort by the participant such as that produced in a controlled experiment, it is hypothesized that throughput can reach the capacity (*i.e.*, the highest theoretically attainable rate) [18]. In the literature, two throughput propositions dominate, namely $1/b$ and the ISO-throughput (Eq. (3)), with no real consensus as to which one is preferable [18].

The capacity in *Variance Model* matches the $1/b$ throughput according to Eq. (9), exactly as advocated by Zhai [41]. Hence, throughput, in the sense of information transmission, is defined as $1/b$. However, such a throughput is likely not what the practitioner is looking for: it characterizes only the transmission part via the second phase of movement,

without accounting for the distance covering part that would include the first phase. To better understand this, one can think of a device, where C is relatively high *i.e.*, variance reduction is very effective, but which creates a lot of variance at the end of the first phase. The result would be a device with high throughput, but an overall poor performance, as measured by MT (and/or T). Unfortunately, there is little that information theory, and thus the *Variance Model*, can offer to provide an all encompassing measure of performance that includes both phases. Hence, we believe that both C and ISO-throughput Eq. 3 are useful metrics. However, we think that ISO-throughput should be renamed to make it clear it is not a throughput related to information transmission — for example the *ISO performance measure*.

8.5 Should participants aim for a 4% Miss Rate?

It has long been argued, based on an information-theoretic formula, that a 4% participant miss rate was something the experimenter should strive for in a Fitts' law experiment [18, 37]. This argument is based on the entropy of a Gaussian distribution, which introduces the 4.133 factor in effective versions of Fitts' law Eq. (2). However, as shown by Eq. (23) τ, that term is arbitrary and actually carries over to the intercept; one could introduce any other value, with the only effect of changing the intercept. In the nominal version of the law, Eq. (22), it even disappears. Hence, there can be no information-theoretic reason to instruct a participant to aim for a 4% miss rate.[15] The dual-minimization protocol solves the issue by considering errors rather than misses (for the difference between misses and errors, refer to [17]).

8.6 Dwell time and stopping conditions

It is generally difficult to infer the movement's end from the trajectories alone because they are characterized by very slow speeds, and potential dwell times (this is in contrast to movement starts). In some contexts, there is no discrete event to signal the end of a movement; as a result, Fitts' law, which relies on a precise determination of movement endpoints, can become unreliable. This difficulty is not present with the method of PVPs: a trajectory with arbitrary dwell can be considered, because what matters is not the actual end of a movement, but the reduction of variance over time.

9 CONCLUSION

In this work, we proposed an alternative to the evaluation of input performance via Fitts' law experiments, by building on our recent theoretical model. We successfully extended the associated method —called the method of PVPs— to deal with 2D data. The method is associated with a dual-minimization protocol which requires a single experimental condition: rather than fully crossing D and W, the researcher should simply select a large D and a small W. Despite this simplification, the method leads to more expressive results, while remaining compatible — and even predicting — Fitts' law parameters. Another advantage of the method is that design choices and analyzes of controlled experiments can be based on the theoretical variance model. Future work will consist in further exploring the method of PVPs. Interesting questions would be to determine the minimum number of samples needed to estimate PVP features reliably, to further reduce the duration of experiments destined to measure input performance. It would also be interesting to study more extreme cases of input performance, such as very low accuracy techniques (*e.g.*, eye/head tracking), or for participants with physical disabilities. Because our method for computing PVPs extends to multiple dimensions, another interesting use case would be the evaluation of 3D pointing. 3D pointing introduces depth perception issues, which, we think, may

[15] It also happens to be very hard to enforce in practice.

lead to slower than exponential decrease of the PVP. On the whole, we believe the method to be a promising alternative to Fitts' law.

ACKNOWLEDGMENTS

We thank Gilles Bailly, as well as the anonymous reviewers, whose advice has helped us greatly improve the paper.

REFERENCES

[1] Johnny Accot and Shumin Zhai. 2003. Refining Fitts' law models for bivariate pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 193–200.

[2] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[3] Bastien Berret, Adrien Conessa, Nicolas Schweighofer, and Etienne Burdet. 2021. Stochastic optimal feedforward-feedback control determines timing and variability of arm movements with or without vision. *PLOS Computational Biology* 17, 6 (2021), e1009047.

[4] S. K. Card, W. K. English, and B. J. Burr. 1978. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics* 21, 8 (1978), 601–613. https://doi.org/10.1080/00140137808931762

[5] E. R. F. W. Crossman. 1957. *The speed and accuracy of simple hand movements*. Technical Report. .

[6] E. R. F. W. Crossman and P. J. Goodeve. 1983. Feedback control of hand-movement and Fitts' law. *The Quarterly Journal of Experimental Psychology* 35, 2 (1983), 251–278.

[7] H. 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)*. ACM, New York, NY, USA, 2813–2822. https://doi.org/10.1145/1753846.1753867

[8] Iddo Eliazar. 2015. How random is a random vector? *Annals of Physics* 363 (2015), 164–184.

[9] Digby Elliott, Steve Hansen, Lawrence EM Grierson, James Lyons, Simon J Bennett, and Spencer J Hayes. 2010. Goal-directed aiming: two components but multiple processes. *Psychological bulletin* 136, 6 (2010), 1023.

[10] Digby Elliott, Werner F Helsen, and Romeo Chua. 2001. A century later: Woodworth's (1899) two-component model of goal-directed aiming. *Psychological bulletin* 127, 3 (2001), 342. https://doi.org/10.1037/0033-2909.127.3.342

[11] Digby Elliott, James Lyons, Spencer J Hayes, James J Burkitt, James W Roberts, Lawrence EM Grierson, Steve Hansen, and Simon J Bennett. 2017. The multiple process model of goal-directed reaching revisited. *Neuroscience & Biobehavioral Reviews* 72 (2017), 95–110. https://doi.org/10.1016/j.neubiorev.2016.11.016

[12] P. 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. https://doi.org/10.1037/h0045689

[13] Paul M Fitts and James R Peterson. 1964. Information capacity of discrete motor responses. *Journal of experimental psychology* 67, 2 (1964), 103. https://doi.org/10.1037/h0045689

[14] K. Gan and E. R. Hoffmann. 1988. Geometrical conditions for ballistic and visually controlled movements. *Ergonomics* 31, 5 (1988), 829–839.

[15] Andrew Gelman and Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

[16] Julien Gori and Olivier Rioul. 2020. A feedback information-theoretic transmission scheme (FITTS) for modeling trajectory variability in aimed movements. *Biological Cybernetics* 114, 6 (2020), 621–641.

[17] J. Gori, O. Rioul, and Y. Guiard. 2017. To Miss is Human: Information-Theoretic Rationale for Target Misses in Fitts' Law. In *CHI'17: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, Denver, USA, 5.

[18] J. Gori, O. Rioul, and Y. Guiard. 2018. Speed-Accuracy Tradeoff: A Formal Information-Theoretic Transmission Scheme (FITTS). *ACM Trans. Comput.-Hum. Interact.* 25, 5, Article 27 (Sept. 2018), 33 pages. https://doi.org/10.1145/3231595

[19] Julien Gori, Olivier Rioul, and Yves Guiard. 2018. Speed-accuracy tradeoff: A formal information-theoretic transmission scheme (fitts). *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 5 (2018), 1–33.

[20] Peter Green and Catriona J. MacLeod. 2016. simr: an R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution* 7, 4 (2016), 493–498. https://doi.org/10.1111/2041-210X.12504

[21] Tovi Grossman and Ravin Balakrishnan. 2005. A probabilistic approach to modeling two-dimensional pointing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 12, 3 (2005), 435–459.

[22] Yves Guiard. 2019. Polar and Cartesian Structure in the Data of Fitts's (1954) Classic Experiments-With a Criterion for Distinguishing a Strong and a Weak Version of Fitts' Law. *Journal of Motor Behavior* , (2019), 1–23.

[23] Y. Guiard and H. B. Olafsdottir. 2011. On the measurement of movement difficulty in the standard approach to Fitts' law. *PloS one* 6, 10 (2011), e24389. https://doi.org/10.1371/journal.pone.0024389

[24] Yves Guiard, Halla B Olafsdottir, and Simon T Perrault. 2011. Fitt's law as an explicit time/error trade-off. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, ACM, New York, NY, USA, 1619–1628. https://doi.org/10.1145/1978942.1979179

[25] Y. Guiard and O. Rioul. 2015. A mathematical description of the speed/accuracy trade-off of aimed movement. In *Proceedings of the 2015 British HCI Conference*. ACM, New York, 91–100.
[26] Peter A Hancock and Karl M Newell. 1985. The movement speed-accuracy relationship in space-time. In *Motor Behavior*. Springer, , 153–188.
[27] E. R. Hoffmann. 2013. Which version/variation of Fitts' law? A critique of information-theory models. *Journal of motor behavior* 45, 3 (2013), 205–215. https://doi.org/10.1080/00222895.2013.778815
[28] Richard J Jagacinski and Donald L Monk. 1985. Fitts' Law in Two dimensions with hand and head movements movements. *Journal of motor behavior* 17, 1 (1985), 77–95.
[29] I. S. MacKenzie. 1989. A note on the information-theoretic basis for Fitts' law. *Journal of motor behavior* 21, 3 (1989), 323–330. https://doi.org/10.1080/00222895.1989.10735486
[30] I. S. Mackenzie. 1992. *Fitts' law as a performance model in human-computer interaction*. Ph. D. Dissertation. University of Toronto.
[31] I. S. MacKenzie and W. Buxton. 1992. Extending Fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, New York, 219–226.
[32] Jörg Müller, Antti Oulasvirta, and Roderick Murray-Smith. 2017. Control theoretic models of pointing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 4 (2017), 27. https://doi.org/10.1145/3121431
[33] Davy Paindaveine. 2008. A canonical definition of shape. *Statistics & probability letters* 78, 14 (2008), 2240–2247.
[34] R. Plamondon and A. M. Alimi. 1997. Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences* 20, 02 (1997), 279–303. https://doi.org/10.1017/S0140525X97001441
[35] Hiroaki Sakoe and Seibi Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing* 26, 1 (1978), 43–49.
[36] R. A. Schmidt, H. Zelaznik, B. Hawkins, James S. F., and J. T. Quinn Jr. 1979. Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychological review* 86, 5 (1979), 415.
[37] R William Soukoreff and I Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International journal of human-computer studies* 61, 6 (2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001
[38] Emanuel Todorov. 2004. Optimality principles in sensorimotor control. *Nature neuroscience* 7, 9 (2004), 907–915.
[39] J. O. Wobbrock, K. Shinohara, and A. Jansen. 2011. The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, ACM, , 1639–1648.
[40] R. S. Woodworth. 1899. Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements* 3, 3 (1899), i.
[41] S. Zhai. 2004. Characterizing computer input with Fitts' law parameters—the information and non-information aspects of pointing. *International Journal of Human-Computer Studies* 61, 6 (2004), 791–809.
[42] S. Zhai, J. Kong, and X. Ren. 2004. Speed–accuracy tradeoff in Fitts' law tasks—on the equivalency of actual and nominal pointing precision. *International journal of human-computer studies* 61, 6 (2004), 823–856.

A EXTRA FIGURES

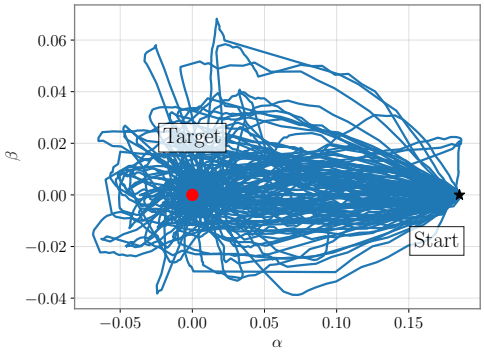

Fig. 8. Left: A set of trajectories from pilot data, projected on the $\{\vec{u}, \vec{v}\}$ basis. We see movement in the direction of the target (α, x-axis) and in the direction orthogonal to it (β, y-axis). Both components are expressed in standard units (m). Close to the target (*i.e.*, in the second phase), the x and y axes seem uncorrelated, indicating a small cross-covariance $\Gamma_{\alpha\beta}$.

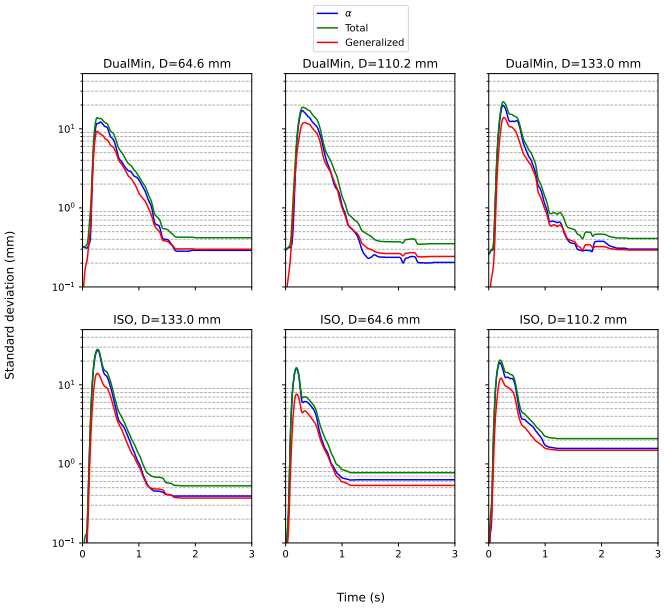

Fig. 9. PVPs for Participant 217235 in the three largest D conditions and for both protocols (*DualMin* top, *ISO* bottom) for the mouse. $\text{PVP}_{\text{total}}$ is always above the two others. PVP_{α} follows $\text{PVP}_{\text{total}}$ early on, indicating $\Gamma_{\alpha\alpha} \gg \Gamma_{\beta\beta}$, but switches to $\text{PVP}_{\text{generalized}}$ towards the very end of the movement, indicating $\Gamma_{\alpha\alpha} \simeq \Gamma_{\beta\beta}$.

B EXTRA TABLES

Table 3. Results (estimated mean and standard deviation for each coefficient) of the multilevel fits for PVP features.

	$\boldsymbol{\tau}$	$\mathbf{D}_{\boldsymbol{\tau}}$	σ_0	C
(Intercept)	0.578^{***}	2.132	4.296^{***}	1.652^{***}
	(0.036)	(1.868)	(0.907)	(0.253)
D	0.003^{***}	0.246^{***}	0.159^{***}	0.004
	(0.000)	(0.021)	(0.010)	(0.002)
W	-0.013^{***}	−0.330	−0.131	0.064^{*}
	(0.003)	(0.265)	(0.128)	(0.027)
devicemouse	-0.221^{***}			3.064^{***}
	(0.035)			(0.136)
devicetouchpad	-0.190^{***}			3.213^{***}
	(0.035)			(0.136)
D × W	0.000	0.003	0.000	−0.001
	(0.000)	(0.004)	(0.002)	(0.000)
D × devicemouse	-0.003^{***}	0.232^{***}	-0.023^{**}	
	(0.000)	(0.015)	(0.007)	
D × devicetouchpad	-0.003^{***}	0.218^{***}	0.089^{***}	
	(0.000)	(0.015)	(0.007)	

Significance: $*** \equiv p < 0.001$; $** \equiv p < 0.01$; $* \equiv p < 0.05$
N = 428, Participant = 12

Table 4. Results (estimated mean and standard deviation for each coefficient) of the linear model fits for the Fitts' law model on data from the *ISO* condition.

	Effective	Nominal
(Intercept)	1.100^{***}	0.708^{***}
	(0.161)	(0.131)
IDe/ID	0.350^{***}	0.402^{***}
	(0.042)	(0.030)
device: mouse/controller	-0.852^{***}	-0.343
	(0.236)	(0.184)
device: touchpad/controller	-0.595^{*}	-0.316
	(0.234)	(0.184)
IDe/ID × device: mouse/controller	-0.162^{**}	-0.252^{***}
	(0.059)	(0.042)
IDe/ID × device: touchpad/controller	-0.136^{*}	-0.165^{***}
	(0.058)	(0.042)
R-squared	0.653	0.765
N	210	210

Significance: $***\equiv p < 0.001$; $**\equiv p < 0.01$; $*\equiv p < 0.05$