Biological
Cybernetics

ORIGINAL ARTICLE

A feedback information-theoretic transmission scheme (FITTS) for modeling trajectory variability in aimed movements

Julien Gori[1] · Olivier Rioul[2]

Abstract

Trajectories in human aimed movements are inherently variable. Using the concept of positional variance profiles, such trajectories are shown to be decomposable into two phases: In a first phase, the variance of the limb position over many trajectories increases rapidly; in a second phase, it then decreases steadily. A new theoretical model, where the aiming task is seen as a Shannon-like communication problem, is developed to describe the second phase: Information is transmitted from a "source" (determined by the position at the end of the first phase) to a "destination" (the movement's end-point) over a "channel" perturbed by Gaussian noise, with the presence of a noiseless feedback link. Information-theoretic considerations show that the positional variance decreases exponentially with a rate equal to the channel capacity C. Two existing datasets for simple pointing tasks are re-analyzed and observations on real data confirm our model. The first phase has constant duration, and C is found constant across instructions and task parameters, which thus characterizes the participant's performance. Our model provides a clear understanding of the speed-accuracy tradeoff in aimed movements: Since the participant's capacity is fixed, a higher prescribed accuracy necessarily requires a longer second phase resulting in an increased overall movement time. The well-known Fitts' law is also recovered using this approach.

Keywords Fitts' law · Speed-accuracy tradeoff · Information theory · Feedback · Variance · Movement · Motor control

1 Introduction

It has long been observed that people routinely adapt their speed to perform aimed movements in a reliable manner: The more accurate a movement, the slower its execution. This so-called speed-accuracy tradeoff has been studied for more than a century by many communities such as experimental psychology (Woodworth 1899; Fitts 1954; Welford 1960; Crossman and Goodeve 1983; Meyer et al. 1988; Crossman 1957), human–computer interaction (HCI) (Card et al. 1978; Soukoreff and MacKenzie 2004), cybernetics (Chan and Childress 1990; Gawthrop et al. 2011), robotics (Sim-

mons and Demiris 2005), and neuroscience (Flanagan and Rao 1995; Khan et al. 2006).

Fitts (1954) provided a simple formula to describe the speed-accuracy tradeoff of simple aimed movements (Fitts' law, see Sect. 2), focusing on the variability of the movement *endpoints*. Fitts' law remains to this day heavily used in various communities, e.g., in HCI to evaluate the performance of input devices (Soukoreff and MacKenzie 2004); unfortunately, it is only a rule of thumb, initially conceived with a vague analogy with information theory (Gori et al. 2018), that only deals with movement endpoints and says little about the entire movement's trajectory.

Since Fitts' seminal 1954 work, many attempts have been pursued to explain Fitts' law as resulting from an underlying mechanism for movement production and control. Examples are neural dynamics models (Bullock and Grossberg 1988), behavioral models (Elliott et al. 2017, 2010), various mathematically formulated models (Plamondon and Alimi 1997; Meyer et al. 1988), and models based on control theory, including simple closed-loop step responses (Müller et al. 2017), deterministic and stochastic optimal control (Todorov and Jordan 2002; Tanaka et al. 2006; Harris and Wolpert

✉ Julien Gori
  juliengori@gmail.com

  Olivier Rioul
  olivier.rioul@telecom-paris.fr

1 LRI, Université Paris-Saclay, CNRS, Inria, 91400 Orsay, France

2 LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France

1998; Flash and Hogan 1985; Guigon et al. 2007; Berret and Jean 2016). This last class of models has been particularly influential in explaining motor planning (Todorov and Jordan 2002). It shows how humans choose one particular path from the infinitely many possible (Flash and Hogan 1985) (this is usually known as the joint-redundancy or degree of freedom problem Rosenbaum 2009; Bullock and Grossberg 1988), and how and when they will correct deviations from that particular path (Todorov and Jordan 2002).

Control theoretic models do suffer from several difficulties. To be operational, they require estimating the impedance of the motor system, i.e., limb and muscles inertia and viscoelastic properties. However, modulating impedance might be a control strategy in its own right (Hogan 1985), and even estimating an impedance that is assumed invariant is not trivial (Buchanan et al. 2004; Müller et al. 2017). Another issue is that it is not clear which cost functions are relevant to the central nervous system (CNS), and how they should be combined (Franklin and Wolpert 2011) in optimal control models.

In this paper, we circumvent these difficulties by using an optimal information transmission model that explains the evolution of variance over time for precise aimed movements, jointly maximizing speed and accuracy without relying on motor impedance. We leverage a computational model rooted in information theory to show that the problem of aiming can be seen as a Shannon-like communication problem (Shannon 1948) with a noisy feedforward channel and a noiseless feedback link. Using Shannon's information theory is not only a reminder of the initial rationale behind Fitts' law (Fitts 1954; Gori et al. 2018); it mostly comes from the observation that noise is "the one key element limiting motor performance" (Franklin and Wolpert 2011)—and that information (communication) theory was conceived with the goal of finding "ways of transmitting the information which are optimal in combating noise" (Shannon 1948). In spite of the claimed differences between our information theoretic and other optimal control schemes, similarities do exist, such as the use of feedback information at the emitter, and the presence of a minimum mean square error (MMSE) estimator at the receiver. Links between control theoretic and information theoretic formalisms have been made (Elia 2004; Tatikonda 2000), and are out of the scope of this paper.

We then introduce *positional variance profiles* (PVPs) as a way to operationalize our model and track the evolution of the variability of trajectories over time. We remark that PVPs are necessarily *unimodal*: A first distance-covering phase where positional variance increases rapidly over time is followed by a much longer second phase—which makes reliable aiming possible. Our model explains this longer second phase of decreasing variance, while an empirical study reveals that the duration of the first phase is approximately constant. We
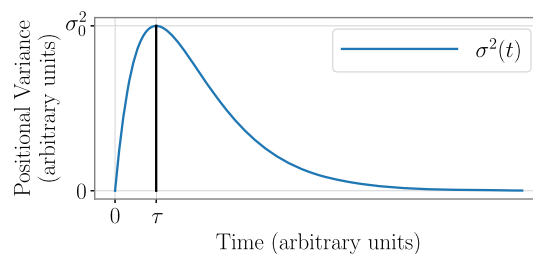

Fig. 1 Ideal two-phase positional variance profile. The transition between the two phases occurs at $(\tau; \sigma_0^2)$

conclude by combining the two phases, which produces a novel derivation for Fitts' law.

The remainder of this paper is organized as follows. Behavioral observations relevant to our model are described in Sect. 2. The information-theoretic model for the second phase is derived in Sect. 3, as the result of a joint optimization of aiming accuracy and movement time. The main theoretical result is that the variance decreases exponentially over time during the second phase. PVPs are described in Sect. 4, and empirical evidence to support our model is provided in Sect. 5. An empirical study of the first phase of the PVPs with increasing variance is presented in Sect. 6. Finally, the synthesis of the two phases leads to the recovery of the well-known Fitts' law in Sect. 7. Section 8 concludes.

2 Background

Our aim in this section is not to present an exhaustive review of motor control models; instead, we present behavioral observations on the variability of movements, the use of feedback information and the continuous and discrete nature of the control of movements that are relevant to our model.

2.1 Variability of movements

Human movement is inherently variable, as recognized since the earliest studies on human produced movement (Woodworth 1899). The variability of the endpoints of movement has been of great interest to behaviorists, leading to our first point of interest, the well-known empirical Fitts' law. A second point of interest is the observation that the evolution of the variance of position over time in aiming tasks has a unimodal profile, of which an "ideal" is represented in Fig. 1. *Fitts law*. Fitts' law describes the time MT it takes to reach a target of size W located D away:

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

$\text{ID} = \log_2 \left(1 + \frac{D}{W}\right)$ is known as the index of difficulty. There is, at present, no way to predict values for a (the intercept, in

seconds) and b (the slope, in seconds per bit), which therefore have to be estimated from empirical data (Soukoreff and MacKenzie 2004; Gori et al. 2018). The inverse of the slope ($1/b$, in bits per second) is often called throughput (Zhai 2004) and is interpreted as a measure of the information-theoretic "capacity" of the human motor system, i.e., the highest rate of information that can be transferred during the task completion.

Fitts' law was first stated by Fitts (1954) as a vague analogy with the Shannon capacity formula (Shannon 1948)—that analogy having been questioned on many occasions (Sheridan and Ferrell 1974; Gori et al. 2018). Due to the informal nature of Fitts' theoretical construct, many variations of the law exist[1] and most of the practices surrounding Fitts' law are empirically guided.

A variation that is used, e.g., in HCI (Soukoreff and MacKenzie 2004; Gori et al. 2018), known as the effective Fitts' law (Zhai et al. 2004), corrects for participants that miss targets over the course of the experiment, replacing W with the actual standard deviation of endpoints σ, estimated from the data:

$$\text{MT} = a + b \log_2 \left(1 + \frac{D}{4.133\sigma}\right) = a + b \, \text{ID}_e \qquad (2)$$

where $\text{ID}_e = \log_2 \left(1 + \frac{D}{4.133\sigma}\right)$ is known as the effective index of difficulty. This formulation is, however, based on the heuristic that an $\varepsilon = 3.88\%$ miss rate should be pursued and lacks a satisfying justification (Gori et al. 2018).

Unimodal positional variance. As noticed in Todorov and Jordan (2002), trajectory variability has "surprisingly received little attention from researchers": Most studies investigating Fitts' law have been limited to the measure of the spread of endpoints (Crossman 1957; Soukoreff and MacKenzie 2004) or single movements (Meyer et al. 1988; Plamondon and Alimi 1997). However, that humans are able to reach almost any target reliably implies that they are able to reduce their endpoint variability at will, which begs the question of how exactly that variability is reduced throughout the trajectory.

The evolution in time of the standard deviation of trajectories from a tapping task was evaluated by Lai et al. (2005) and appeared *unimodal* (an increasing phase was followed by a decreasing phase). Other studies (Khan et al. 2006; Van der Meulen et al. 1990) have represented positional variance at specific kinematic markers (peak acceleration, peak velocity, peak deceleration, movement time), which also suggests unimodal variance profiles. Finally, Gutman and Gottlieb (1992) and Gutman et al. (1993) proposed a feedforward model where a primitive trajectory is rescaled in time and space depending on the task. They found their model predicted unimodal variance profiles, in line with empirical observa-

tions.[2] Thus, evidence suggests that the positional variance has a unimodal profile in time; this was also verified in our empirical analysis (see Sect. 6).

The model that we propose accounts for the variance decreasing phase of movement: We namely show that the variance, in the decreasing phase, does so at a characteristic exponential rate.

2.2 Feedforward and feedback information

It is well known that humans cannot function properly without feedback, see, e.g., Wiener's account (Wiener 1961, p. 95) on two patients suffering from a lack of voluntary muscle movement coordination known as *ataxia* caused by a lack of proprioceptive feedback information at the brain. Another illustration comes from the fact that older adults have a diminished sense of proprioception which seems to account for a decrease in their performance compared with younger adults (Ketcham and Stelmach 2004). Movement generation also relies on visual feedback mechanisms: Performing precise movements with closed eyes is near impossible, and various experiments on occlusion and removal of light (Elliott et al. 1995; Zelaznik et al. 1983) or removal of cursor (Chua and Elliott 1993) confirm an effect of visual feedback on virtually all kinematic properties, including accuracy and movement time, on movements as short as 100 ms (Elliott et al. 2010; Carlton 1992; Zelaznik et al. 1983).

However, many works also indicate feedforward control (Mehta and Schaal 2002), in particular those with deafferented monkeys (Polit and Bizzi 1978, 1979; Bizzi et al. 1984). Incorporating both feedforward and feedback mechanisms is thus necessary when designing a scheme to explain aimed movement production (Wolpert 1997). This was already recognized in earliest descriptions of aimed movements by Woodworth (1899), who hypothesized that an aimed movement is constituted of two serial components: An initial adjustment, whose main purpose is to cover distance, followed by a homing-in phase that relies on vision to ensure accuracy. This two-component model has been recently refined under the impulse of Elliott et al. (2001, 2010, 2017): A first *planned* component gets the limb close to the target area. This component is based on internal representations and is associated with a velocity regulation through proprioceptive feedback. Typically high speeds and low jerk trajectories are associated with this component. When time permits, a second corrective portion is engaged to reduce spatial discrepancy between limb and target. This process, highly dependent on foveal (central) vision, involves com-

---

[1] For example, at least a dozen different formulations for ID exist (Plamondon and Alimi 1997), including Fitts' original ID $= \log_2 \left(\frac{2D}{W}\right)$.

[2] They did find bimodal profiles, and so did Darling and Cooke (1987) in another study, on fast elbow only (single-joint) flexions. This differs from our work which tackles multi-joint movements.

puting the difference between limb and target position and issuing discrete corrections.

Our model incorporates feedback and feedforward mechanisms via a feedback and a feedforward channel. We will also find many resemblances between the two aforementioned phases of variance and the two components of Woodworth and Elliott.

2.3 Continuous, discrete, and intermittent control

Woodworth (1899) was the first to suggest that aiming was comprised of a first distance-covering phase followed by a second homing-in phase, but the segmentation was never explicitly performed, including in its modern exposition (Elliott et al. 2001, 2010). It was later observed that the velocity during movement vanished multiple times before its end, which was interpreted as the transition from one submovement to another—the whole movement being composed of concatenated submovements—which implies a discrete type of control (Crossman and Goodeve 1983; Keele 1968; Meyer et al. 1988). Several models have used this idea of submovements, including Crossman and Goodeve's (1983) deterministic iterative control model, and its stochastic counterpart due to Meyer et al. (1988), which have proved quite popular.

There are, however, many reports of improvement of overall performance metrics in the presence of feedback, even though no distinct changes occur in the kinematic profiles (Pélisson et al. 1986; Desmurget and Grafton 2000; Elliott et al. 1991). This would indicate continuous control, and some successful optimal control models are indeed continuous. Yet another form of control is intermittent control, where, e.g., observations are performed continuously but actions are taken intermittently (Gawthrop et al. 2011).

It has been suggested that training and learning leads to the transition from discrete or intermittent control to continuous control (Woodworth 1899; Proteau et al. 1987; Elliott et al. 2010). It has also been suggested that intermittent control can appear to be continuous—so-called masquerading (Gawthrop et al. 2011). This can lead to difficulties with models that form predictions that are based on intermittent/discrete style control strategies, e.g., if the quantity (e.g. amplitude, frequency) or quality of submovements are predicted. Our use of variance profiles, that relies on statistical averaging, allows us to form predictions regardless of the actual type of control.

3 Variance-decreasing phase: Theoretical model

Contrary to most models of movement, where initial position is given (e.g., the optimal control theoretic framework

(Todorov and Jordan 2002; Tanaka et al. 2006; Harris and Wolpert 1998; Flash and Hogan 1985; Guigon et al. 2007; Berret and Jean 2016), the input to our model is a *random* variable, corresponding to the position when positional variance is maximized (time $t = \tau$ in Fig. 1). In this section, some theorems and lemmas are given. All full proofs are delayed to Appendix A.3.

3.1 Notations

- bold letters $\mathbf{X}$ and $\mathbf{X}_i$ denote random variables which can be indexed by i in the list $\mathbf{X}^n = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_i, \ldots, \mathbf{X}_n)$;
- $\mathcal{N}(\mu, \sigma^2)$ is the μ-centered Gaussian distribution with variance σ^2;
- $\mathbb{E}[\mathbf{X}]$ and $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$ are, respectively, the mathematical expectation of $\mathbf{X}$ and the conditional expectation of $\mathbf{X}$ given $\mathbf{Y}$;
- $H(\mathbf{X})$ and $H(\mathbf{X}|\mathbf{Y})$ are, respectively, the differential entropy and the differential conditional entropy, defined by $H(\mathbf{X}) = -\mathbb{E}[\log_2 p(\mathbf{X})]$ and $H(\mathbf{X}|\mathbf{Y}) = -\mathbb{E}[\log_2 p(\mathbf{X}|\mathbf{Y})]$ and are expressed in bits, were $p(\mathbf{X})$ and $p(\mathbf{X}|\mathbf{Y})$ are the probability density functions (pdf) of $\mathbf{X}$ and $\mathbf{X}|\mathbf{Y}$. An important example is the Gaussian entropy (Cover and Thomas 2012): If $\mathbf{X} \sim \mathcal{N}(0, \sigma^2)$, then $H(\mathbf{X}) = \frac{1}{2} \log_2(2\pi e \sigma^2)$.

3.2 Information-theoretic model description

Our goal is to explain the reliability of human-aimed movements, i.e., how the inherent variability of a set of trajectories gets reduced over time (from τ, the time of the maximum of the PVP, onward to the end of the movement). In contrast, existing models of goal-directed movements usually predict trajectories formed between and initial position and final position.

The position of the limb extremity ('limb' in short) at the end of the first phase is *random*: The distance to the target is modeled by the random variable $\mathbf{A}$.

Thanks to visual and proprioceptive feedback information the limb position is known at the brain level. Due to eye–hand coordination and fast eye dynamics, the eye is usually pointing toward the target long before the end of the movement (Elliott et al. 2010). Hence, we assume that $\mathbf{A}$ is known at the brain level and can be easily evaluated.[3]

We take it that $\mathbf{A}$ is a centered Gaussian with variance σ_0^2: $\mathbf{A} \sim \mathcal{N}(0, \sigma_0^2)$. We consider $\mathbf{A}$ to be centered, since movements can undershoot and overshoot the target. In general, movements can come from any direction (left or right in this 1D model), hence an overshoot for one is an undershoot for

---

[3] In fact, it can be estimated by the eye if the limb is close enough to the target. This is indeed the case since most of the distance to the target has been covered during the first phase, see Sect. 6.
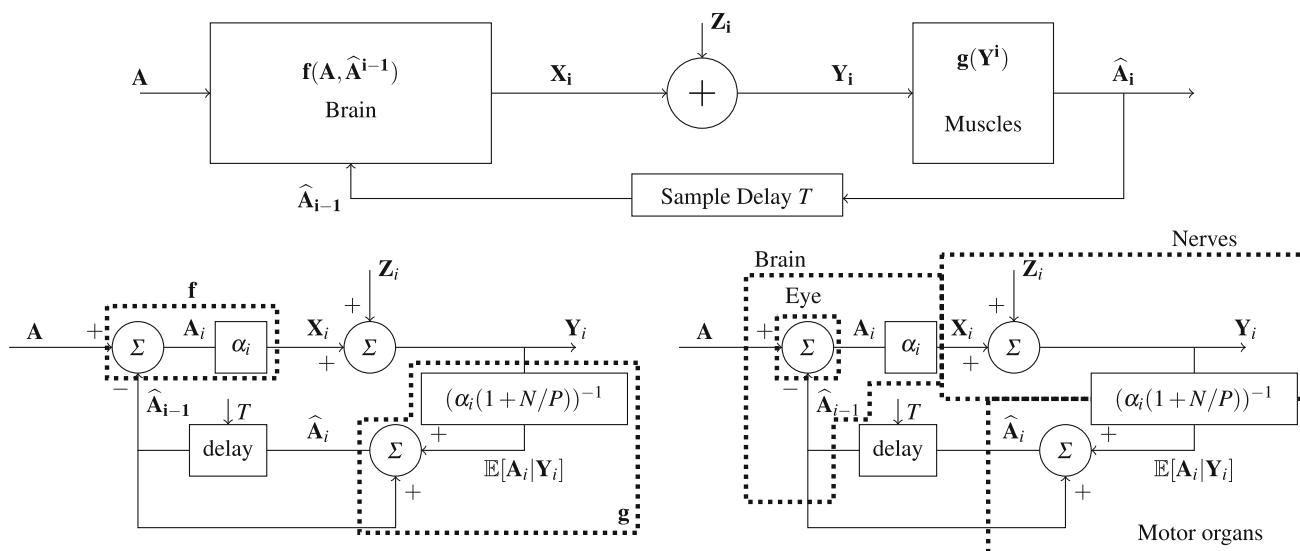
Fig. 2 Top panel: Information-theoretic model for the aiming task. Bottom left panel: Implementation of the model. Bottom right panel: Physiological representation

the other, symmetrizing $\mathbf{A}$. Empirical evidence in Sect. 5 will further support the Gaussian hypothesis. The first iteration of the scheme is now explained.

First iteration of the scheme. The second phase of movement is commensurate to sending $\mathbf{A}$, the remaining algebraic distance to the target, from the brain to the limb (to be read with the top panel of Fig. 2):

– Given $\mathbf{A}$, the brain outputs an amplitude $\mathbf{X}_1$ to be sent to the limb:

$$\mathbf{X}_1 = \mathbf{f}(\mathbf{A}), \tag{3}$$

where $\mathbf{f}$ is a *deterministic*, yet unknown function performed by the brain, known as the *encoder* in the vocabulary of information theory.

– The variability of the human motor system is accounted for by a noisy (Gaussian) transmission from brain to limb. The output of the channel $\mathbf{Y}_1$ reads

$$\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{Z}_1, \tag{4}$$

where $\mathbf{Z}_1 \sim \mathcal{N}(0, N)$ is the "noise." The noise is taken as additive because the remaining distance to the target is small; hence, the noise is largely independent from $\mathbf{X}_1$, see Todorov (1998). This transmission model is the well-known Additive White Gaussian Noise (AWGN) channel (Shannon 1948).

– The actual distance covered by the limb $\widehat{\mathbf{A}}_1$ is the result of some yet unknown function g (the *decoder* in the vocabulary of information theory), applied by the motor organs

to the received $\mathbf{Y}_1$:

$$\widehat{\mathbf{A}}_1 = g(\mathbf{Y}_1). \tag{5}$$

– $\widehat{\mathbf{A}}_1$ is returned to the brain *via* noiseless[4] feedback (e.g., visual, proprioceptive) and used together with $\mathbf{A}$ to produce a new input: $\mathbf{X}_2 = f(\mathbf{A}, \widehat{\mathbf{A}}_1)$—we later show that the best strategy is to use a comparison between $\mathbf{A}$ and $\widehat{\mathbf{A}}_1$ as the subsequent signal.

The scheme then progresses iteratively, each step having a constant duration of T seconds. This puts the limiting factor for time not on the dynamics of moving the limb, but on the delay associated with the feedback transmission.[5]

General form of the iterative scheme. At iteration $i \in \{1, \ldots, n\}$, the scheme is described by the following equations (see the top panel of Fig. 2):

1. The encoder ($\mathbf{f}$) produces $\mathbf{X}_i$ from $\mathbf{A}$ and all received feedback information $\widehat{\mathbf{A}}^{i-1} = \{\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2, \ldots, \widehat{\mathbf{A}}_{i-1}\}$:

$$\mathbf{X}_i = f(\mathbf{A}, \widehat{\mathbf{A}}^{i-1}). \tag{6}$$

---

[4] We assume that in stereotypical controlled experimental tasks such as those used in this paper, there are no perception issues. A pathology could however be simulated by e.g., introducing on purpose a strong deterioration of the quality of the feedback link.

[5] It is equivalent to consider that all components have zero-delay except for the feedback link, or that the delay is spread over all components. For example, the delay taken into account for the feedback link encompasses the delay associated with the feedforward transmission.

2. The AWGN channel:

$$\mathbf{Y}_i = \mathbf{X}_i + \mathbf{Z}_i. \tag{7}$$

3. The output ($\widehat{\mathbf{A}}_i$) of the decoder ($\mathbf{g}$) is a function of all received channel outputs $\mathbf{Y}^i = \{\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_i\}$ thus far:

$$\widehat{\mathbf{A}}_i = \mathbf{g}(\mathbf{Y}^i). \tag{8}$$

The functions $\mathbf{f}$ and $\mathbf{g}$ are assumed to be deterministic and are causal by definition; their exact form, however, has yet to be specified. To actually use all previously received signals since the start of the movement, $\mathbf{f}$ and $\mathbf{g}$ require access to some form of memory. Although this is plausible at the brain level for $\mathbf{f}$ (6), it seems less so at the limb level for $\mathbf{g}$ (8). We next determine $\mathbf{f}$ and $\mathbf{g}$ by solving an optimization problem, and we show that no memory is actually required: The optimal solution—even when assuming the possibility of memorization—is memoryless.

In Shannon's communication-theoretic terms, the aiming task in the second phase can be seen as the transmission of a real value (distance from target at the end of the first phase) from a "source" (CNS) to a "destination" (limb extremity) over a noisy forward Gaussian channel with noiseless feedback. In human-centered terms, that phase makes sure that the limb *reliably* reaches the target, once most of the distance has been covered.

3.3 Bounds on transmitted information

We now leverage information-theoretic definitions.

– $P_i = \mathbb{E}[\mathbf{X}_i^2]$ is the average power (variance) of the channel input at iteration i.
– The quadratic *distortion* $D_i = \mathbb{E}[(\mathbf{A} - \widehat{\mathbf{A}}_i)^2]$ is the mean-squared error of the estimation of $\mathbf{A}$ by $\widehat{\mathbf{A}}_i$ after i iterations.
– $I(\mathbf{A}; \widehat{\mathbf{A}}_i) = H(\mathbf{A}) - H(\mathbf{A}|\widehat{\mathbf{A}}_i)$ is Shannon's mutual (transmitted) information (Cover and Thomas 2012) between $\mathbf{A}$ and $\widehat{\mathbf{A}}_i$.
– Shannon's capacity C (Shannon 1948; Cover and Thomas 2012) of the forward AWGN Channel (for one channel use) under the power constraint $P_i \leq P$ and with noise power N is

$$C = \frac{1}{2} \log_2(1 + P/N) \text{ (bit per channel use).} \tag{9}$$

Shannon's (1948) capacity formula is a corollary to the channel coding theorem (for AWGN channels), which expresses a compromise between a certain measure of speed (rate of information) and a certain measure of accuracy (reliability of a transmission). Goal-directed movements, when

modeled according to Fig. 2, are shown to entail a similar communication tradeoff through the following theorem.

Theorem 1 *Consider the noisy transmission scheme with noiseless feedback of Fig. 2. For a zero-mean Gaussian source* $\mathbf{A}$ *with variance* σ_0^2, *we have after n iterations*

$$\frac{1}{2} \log \frac{\sigma_0^2}{D_n} \underset{(a)}{\leq} I(\mathbf{A}, \widehat{\mathbf{A}}_n) \underset{(b)}{\leq} nC. \tag{10}$$

The inequality on the left expresses the minimum amount of information that needs to be transmitted from the brain to the limb to reduce the positional variability from the initial variance (σ_0^2) to the variance after n iterations (D_n). The inequality on the right expresses the maximum amount of information that can be transmitted over the n times fully exploited noisy channel (nC). Since the transmitted information per channel use can never exceed C, and since being more accurate requires sending larger amounts of information, more iterations of the scheme (hence more time) are needed for more precise tasks—in line with the speed-accuracy tradeoff.

For a given channel capacity C and a given number n of iterations of the scheme, maximizing accuracy is equivalent to minimizing the trajectory variance D_n. Similarly, for a given accuracy D_n, minimizing time is equivalent to minimizing n. Optimal aiming, which we define by those movements that achieve the best possible accuracy in the least amount of time is thus achieved when equality holds in (10), i.e., when:

$$\frac{1}{2} \log \frac{\sigma_0^2}{D_n} = I(\mathbf{A}, \widehat{\mathbf{A}}_n) = nC. \tag{11}$$

The next subsection determines the conditions under which this maximal exploitation of the channel is actually reached.

3.4 Optimal aiming (achieving capacity)

Lemma 1 *Optimal aiming can be achieved if, and only if, the following conditions hold:*

1. *all considered random variables* $\mathbf{A}$, $\widehat{\mathbf{A}}^i$, $\mathbf{A} - \widehat{\mathbf{A}}^i$, $\mathbf{X}^i$, $\mathbf{Y}^i$, $\mathbf{Z}^i$ *are Gaussian;*
2. *all input powers are equal (to, say, P):* $P_i = \mathbb{E}[\mathbf{X}_i^2] = P, \quad \forall i;$
3. *endpoints* $\widehat{\mathbf{A}}^i$ *are mutually independent;*
4. *channel outputs* $\mathbf{Y}^i$ *and errors* $\mathbf{A} - \widehat{\mathbf{A}}^i$ *are independent;*
5. $\widehat{\mathbf{A}}_i = \mathbf{g}(\mathbf{Y}^i)$ *is a sufficient statistic of* $\mathbf{Y}^i$ *for* $\mathbf{A}$.

Optimal aiming can only be achieved if all the variables are Gaussian (see Sect. 5 for an empirical investigation). This Gaussianity suggests that $\mathbf{f}$ and $\mathbf{g}$ are linear functions, as the

family of Gaussian distributions is closed under linear combinations. For Gaussian variables, independence is equivalent to decorrelation (or orthogonality), a property we take advantage of in the proofs.

Orthogonality between channel outputs and errors, and between updates (Conditions 3. and 4. of Lemma 1) suggest MMSE estimation (similarly to the Kalman filter 1980) heavily used in the stochastic optimal control setting). Bridging this result with the previous, it is well-known that the MMSE estimator—which minimizes D_n—is a linear function in the Gaussian setting. The general structure of the scheme is thus expected to be linear, with an MMSE estimator on the receiver side.

Since $\mathbf{g}(\mathbf{Y}^i)$ is a sufficient statistic of $\mathbf{Y}_i$ for $\mathbf{A}$, it does not matter if the feedback comes from the endpoints $\widehat{\mathbf{A}}^i$ or from the outputs of the channel $\mathbf{Y}^i$. This is particularly relevant in the present context, since visual and proprioceptive motor feedback information can originate from several sources in the human body.

We now assume for the remainder of the paper that movements are optimal, i.e., conditions of Lemma 1 hold. By further working out the conditions of Lemma 1, we can derive the structure of $\mathbf{f}$ and $\mathbf{g}$. We first obtain $\mathbf{g}$ by using the well-known *orthogonality principle*: If $\mathbf{A}$ is to be estimated from the observed data $\mathbf{Y}^i$ by the unbiased estimator $\hat{\mathbf{A}}(\mathbf{Y}^i)$ then the following statements are equivalent:

- $\hat{\mathbf{A}}(\mathbf{Y}^i) = \mathbb{E}[\mathbf{A}|\mathbf{Y}^i] = \mathbb{E}[\mathbf{A}\mathbf{Y}^i]^t \mathbb{E}[\mathbf{Y}^i(\mathbf{Y}^i)^t]^{-1} \mathbf{Y}^i$ is the MMSE estimator;
- $\mathbb{E}\left[(\mathbf{A} - \hat{\mathbf{A}}(\mathbf{Y}^i))\mathbf{Y}_i\right] = 0, \ \forall i$.

From condition (4), we have that $\mathbb{E}[(\mathbf{A} - \mathbf{g}(\mathbf{Y}^i))\mathbf{Y}_i] = 0, \ \forall i$, hence, the following theorem results from a direct application of the orthogonality principle.

Theorem 2 *For the optimal transmission scheme, $\mathbf{g}(\mathbf{Y}^i)$ is the MMSE estimator:* $\mathbf{g}(\mathbf{Y}^i) = \mathbb{E}\left[\mathbf{A}|\mathbf{Y}^i\right]$.

The optimal scheme thus yields an endpoint $\widehat{\mathbf{A}}_i = \mathbf{g}(\mathbf{Y}^i)$ obtained as the best least-squares estimation of $\mathbf{A}$ from all the current observations of the (independent) channel outputs $\mathbf{Y}^i = (\mathbf{Y}_1, \ldots, \mathbf{Y}_i)$.

Theorem 3 *For the optimal transmission scheme, $\mathbf{f}$ produces a scaled version of the estimation error:*

$$\mathbf{X}_i = \mathbf{f}(\widehat{\mathbf{A}}^{i-1}, \mathbf{A}) = \alpha_i (\mathbf{A} - \widehat{\mathbf{A}}_{i-1}) \tag{12}$$

$$= \alpha_i (\mathbf{A} - \mathbb{E}\left[\mathbf{A}|\mathbf{Y}^{i-1}\right]) = \alpha_i (\mathbf{A} - \mathbf{g}(\mathbf{Y}^{i-1})), \tag{13}$$

where α_i is such that the power constraint $\mathbb{E}[\mathbf{X}_i^2] = P$ is met.

The signal sent at the input of the channel is thus simply the difference between the initial *message* $\mathbf{A}$ and its most recent estimate $\widehat{\mathbf{A}}_{i-1}$, rescaled to meet the power constraint. The previous two theorems formally define the encoding function $\mathbf{f}$ and decoding function $\mathbf{g}$.

The encoding function $\mathbf{f}$ is mathematically simple and appears biologically feasible, since the difference between $\mathbf{A} - \widehat{\mathbf{A}}_{i-1}$ is simply the remaining distance to the target, which can be estimated easily by the eye. In addition, scaling processes (for the α_i's) have been identified in the literature, e.g., within the basal ganglia (Rosenbaum 2009). The decoding function $\mathbf{g}$ is expressed as a function of $\mathbf{Y}^i$, which suggests that the motor organs memorize all channel outputs for later use. To our knowledge, this seems unreasonable. Fortunately, as we next show, there is no need for a memory within the motor organs.

Theorem 4 *Let $\mathbf{A}_i = \mathbf{X}_i/\alpha_i$ be the unscaled version of $\mathbf{X}_i$, with $\mathbf{A}_1 = \mathbf{A}$ and $\alpha_1 = 1$. We have:*

$$\mathbb{E}[\mathbf{A}|\mathbf{Y}^i] = \sum_{j=1}^{i} \mathbb{E}[\mathbf{A}|\mathbf{Y}_j] = \sum_{j=1}^{i} \mathbb{E}[\mathbf{A}_j|\mathbf{Y}_j] \tag{14}$$

$$= \sum_{j=1}^{i} \frac{1}{\alpha_i} (1 + N/P)^{-1} \mathbf{Y}_j \tag{15}$$

The theorem, through (14), shows that the decoding process is recursive: At each step, a "message" $\mathbf{A}_i$ that is independent from the previous ones ($\{\mathbf{A}_1, \ldots, \mathbf{A}_{i-1}\}$) is formed, sent to the channel and estimated optimally by least-square minimization. Say A is initially sent over the channel: The first estimate at the decoder is $\mathbb{E}[\mathbf{A}|\mathbf{Y}_1]$; the motor organs respond, and the limb moves by $\widehat{\mathbf{A}}_1 = \mathbb{E}[\mathbf{A}|\mathbf{Y}_1]$. The remaining distance is $\mathbf{A}_2 = \mathbf{A} - \mathbb{E}[\mathbf{A}|\mathbf{Y}_1]$, and the next estimate is $\mathbb{E}[\mathbf{A}_2|\mathbf{Y}_2]$. The limb moves again, by $\widehat{\mathbf{A}}_2 = \mathbb{E}[\mathbf{A}_2|\mathbf{Y}_2]$. The remaining distance is now $\mathbf{A}_3 = \mathbf{A} - (\mathbb{E}[\mathbf{A}|\mathbf{Y}_1] + \mathbb{E}[\mathbf{A}_2|\mathbf{Y}_2])$. The estimate at each iteration is thus simply the sum of a new estimate and the sum of all previous estimates, i.e., exactly the total distance covered since $\mathbf{A}$. This distance, and hence the channel outputs, do not need to be stored by some internal memory inside the motor organs—they are in fact "memorized" by the limb simply remaining where it is in the absence of a signal.[6] The optimal procedure is thus incremental and optimal at each step and is achieved on-line without memory.

For completeness, we finally check optimality in (11) by evaluating the distortion; we also provide the closed form expression for α_i.

---

[6] Note that we assume that the dynamics of the limbs are fast enough with respect to T that the given distance can always be covered; this reflects one of the main ideas behind this work, namely that for the homing-in phase, which is exceptionally long compared with the distance covered, the limb impedance is not the limiting factor.

Theorem 5 *The quadratic distortion $D_i = \mathbb{E}[(\mathbf{A} - \widehat{\mathbf{A}}_i)^2]$ decreases geometrically in the number of iterations i:*

$$D_i = \frac{\sigma_0^2}{(1 + P/N)^i}. \tag{16}$$

The scaling factor α_i increases geometrically in the number of iterations i:

$$\alpha_i = \alpha_0 (1 + P/N)^{i/2}, \tag{17}$$

where $\alpha_0 = \frac{\sqrt{P}}{\sigma_0}$.

The capacity C is exactly achieved and the distortion decreases geometrically (divided by $1 + P/N$) at each iteration. An implementation of the scheme, based on the equations given in the theorems, is presented in the bottom left panel of Fig. 2, and a tentative mapping to physiological components is given in the bottom right panel of Fig. 2.

The formulas for the distortion and the general scheme have been previously obtained in an information-theoretic context by Gallager and Nakiboglu (2010) who discussed an older scheme by Elias (1957). To our knowledge, the constructive proofs of the Elias scheme given here, as well as its application to model goal-directed movements, are novel.

3.5 From a discrete time model to continuous PVPs

We presented a discrete-time scheme in line with the ideas of iterative corrections. If the iteration time is constant and equal to T and n, iterations have been completed, then the duration of the second phase t is given by $t = nT$. The profile of variance is then a discrete set of points at time $\{t_k = kT\}_{k=1}^n$, where variance amplitudes are given by Theorem 5.

When we operationalize the model through the monitoring of PVPs, we note:

– That T is constant might be true only on average. Different participants may have different values of T. Given a single participant, T may also vary with fatigue, learning, etc.
– Even if T were constant, the uncertainty associated with determining the starting time (and as a result τ) would induce iterations that are not synchronized.

The iterative corrections are thus in practice desynchronized. As a result, when considering sufficiently many trajectories during the construction of a PVP, it is likely that any time interval, however, small, will contain at least one new correction from one of the trajectories. Asymptotically (i.e., considering infinitely many trajectories), we can therefore assume an infinitesimal feedback time δT, leading to a continuous time formulation. This also means the number

of iterations $n = t/\delta T$ goes to infinity. This does not mean that each trajectory is obtained as the result of a continuous on-line control; we posit that a set of trajectories can be described by a continuous model where variance decreases smoothly over time.

With $n = t/\delta T$ one can rewrite (11) as

$$t = \frac{1}{C'} \log_2 \frac{\sigma_0}{\sigma(t)} \tag{18}$$

where the sample mean squared error (continuous time) $\sigma^2(t)$ is taken for the ensemble mean squared error (discrete time) D_n and $C' = C/\delta T$ is the capacity in *bit per second*. Note that this local formulation holds for any time interval in the second phase: For arbitrary $\Delta t \geq 0$

$$\log_2 \sigma(t + \Delta t) = \log_2 \sigma(t) - C' \Delta t, \tag{19}$$

as long as t and $t + \Delta t$ are in the second phase. In the remainder of this work, we do not discriminate between C' and C and use invariably C for simpler notations.

The remainder of this work is almost entirely empirical: We first validate the theoretical results just presented on the variance decreasing phase. Then, we propose an empirical study of the first phase, since no useful information-theoretic scheme can predict an increasing variance. Fitts' law is then derived by combining the results obtained on both phases.

4 Method

4.1 Dataset presentation

Many studies of goal-directed movements having previously been conducted, we have re-analyzed existing datasets to provide empirical support for our model.

The first is from a study by Guiard et al. (2011), following a discrete protocol (Fitts and Peterson 1964). The task of the participant was to move a cursor from a given starting position to a given line $D = 150\,\text{mm}$ away, following 5 different instructions: #1 maximize speed (U-Fast), #2 emphasize speed (Fast), #3 balance speed and accuracy (Balanced), #4 emphasize accuracy (Precise), #5 maximize accuracy (U-Precise).

The second is *pointing dynamics dataset* (PD-dataset) by Müller et al. (2017). The experiment is a replication of Fitts' well-known, one-dimensional 1954 reciprocal task (Fitts 1954), where input is tracked by a mouse. Participants have to move back and forth as fast as possible between two targets of same width W and located D apart. D (in m) $\in \{0.212, 0.353\}$ and ID $\in \{2, 4, 6, 8\}$ [see Eq. (1)] were fully crossed. The PD-dataset is already pre-processed, regularly sampled and denoised with a low-pass filter at about

8 Hz. We performed a similar interpolation and low-pass filtering process on the G-dataset. Full details are available in the original publications.

In both datasets, position is a one-dimensional real signal. As explained below, the two paradigms provide a general picture of the speed-accuracy tradeoff in precise aimed movements:

– *Manipulating speed and accuracy.* For the PD-dataset, accuracy is an independent variable, manipulated via a visual target of size W. In the G-dataset, the participants do not aim toward a target but to a point in space. They are in charge of balancing the speed and accuracy of movements to conform to the instruction given by the experimenter: Both movement time and accuracy are thus dependent measures.
– *Discrete or reciprocal task.* The G-dataset was acquired using a discrete task (Fitts and Peterson 1964), where the cursor is repositioned at the start after each movement, whereas the PD-dataset was acquired using a reciprocal task (Fitts 1954).
– *Multi-joint movements.* In both experiments, the movements solicited several joints: Fingers, wrist, elbow, shoulder, and sometimes even elicited movement from the back. These should be differentiated from single-joint movements (e.g., wrist-only movements Meyer et al. 1988) because the latter eliminate the redundancy of the degrees of freedom, i.e., the issue at the heart of variability in aimed movements (Guigon et al. 2008; Todorov and Jordan 2002).

4.2 Positional variance profiles (PVPs)

Tracking the evolution of variance over time is achieved via so-called positional variance profiles (PVP). The first step in computing PVPs is to identify individual movements from the time series of the position. This is achieved using a home-made segmentation algorithm, described in Appendix A.1.

PVPs are then computed, considering all the trajectories produced for a given condition, using following operations:

1. The trajectories are synchronized by using the starting time as the new time origin.
2. Each trajectory is extended by padding with the final value of position so that all movements last the same amount of time (say 2 s). Indeed, while movements may stop at some point, the position signal remains constant as long as there is no movement, just as if the participant paused longer before going on to the next movement[7];

---

[7] There are thus practically 3 phases; the 2 phases discussed previously, and a third stationary phase, where position is conserved and nothing happens.
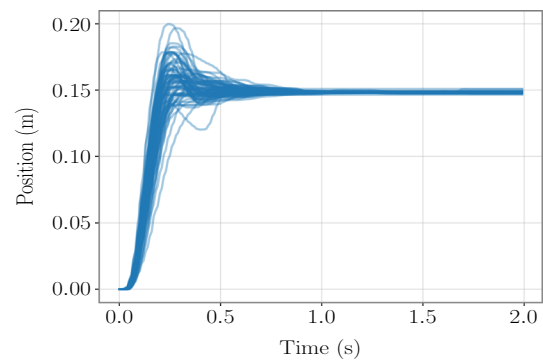

Fig. 3 The set of trajectory for Participant X performing under condition #3 (Balanced) for the G-dataset. All trajectories have been time-shifted so as to start at the origin; after about 1 s, the trajectories are all stationary

3. Once the set of trajectories has been synchronized and extended, compute the variance of the position for this set. This time-series representing the evolution of variance over time is what we call a PVP.

Figure 3 displays a set of synchronized trajectories for a participant of the G-dataset, and Fig. 4 displays several example PVPs computed from empirical data. We asserted the unimodality of PVPs for both datasets. 94% (75 out of 80) of PVPs were found unimodal in the G-dataset; 92% (88 out of 96) in the PD-dataset. Non-unimodal profiles had small secondary peaks due to trajectory outliers. Hence, as expected from Sect. 2, variance profiles for both datasets can be considered unimodal.

4.3 Calculations performed

For each dataset, we segmented trajectories as described in Sect. 4. MT was computed (black diamonds on, respectively, the top and bottom left panel of Fig. 6), and Fitts' law evaluated (results deferred to Sect. 7), to verify that our segmentation procedure produced results in accordance with what is expected from the literature (Soukoreff and MacKenzie 2004).

We then computed the PVPs for each block. The decreasing and increasing phases of variances were straightforwardly identified by locating the mode (τ, σ_0^2) of the PVP. To evaluate our model, the decreasing phase of variance was fitted with a spline, as described in Sect. 5. To study the increasing phase of variance, we systematically observed the effect of the different experimental conditions on σ_0 and τ, as well as D_τ, the distance covered at $t = \tau$, since our theoretical model in Sect. 3 builds heavily on the position signal.

A box-plot representation of τ (See Appendix A.2) with data grouped per participant revealed a large disparity in the

average τ per participant (ranging from $\tau = 0.2$ for P2 to $\tau = 0.45$ for P9 in the PD-Dataset). Similarly, the value of C, the capacity in our theoretical model, is by definition dependent on each participant. Hence, we fitter mixed linear model with random intercepts clustered by participant in our regression models when analyzing the PD-dataset and used one-way repeated measures (RM) ANOVA when analyzing the G-dataset. Greenhouse–Geisser correction, indicated by the $_{GG}$ subscript, was applied when the sphericity assumption was broken.

5 Exponentially decreasing variance phase: Empirical validation

Theorem 5 and its continuous counterpart (19) forecast a decrease in the positional standard deviation at an exponential rate C. The parameter C is a constant that characterizes the channel, is inherent to each participant and should therefore be unrelated to external constraints, such as task geometry. This makes for the following predictions:

Prediction 1 The rate at which the positional standard deviation decreases is exponential.

Prediction 2 That exponential rate C is constant and does not depend (within limits) on the task parameters (i.e., D, W, ID, speed-accuracy strategies).

Both predictions are tested on the PD and G-datasets. It is instructive to look at empirical PVPs in log-lin scale before any statistical analysis, see Fig. 4; the left panel represents the 8 PVPs for a single participant of the PD-dataset, performing all 8 conditions, the middle panel represents the 5 PVPs for a single participant of the G-dataset, performing all 5 conditions, and the right panel is a single PVP taken from the G-dataset with the corresponding spline fit. One sees that, as expected from (19), the second phase is close to linear and that the slopes are all more or less confounded. This suggests that values of C for each PVP are about equal, suggesting that Predictions 1 and 2 hold.

5.1 Prediction 1: Exponential decrease in standard deviation

As shown in (19), an exponential decrease in the standard deviation will appear linear in a log-lin scale. Hence, in a log-lin scale, the PVP is theoretically composed of the following three phases:

1. a first phase of duration τ;
2. a second phase which decreases linearly, from standard deviation σ_0 until some value, say σ_∞;

3. a third stationary phase where standard deviation is constant and equal to σ_∞, which lasts as long as the duration of the extended trajectories used to construct the PVP.

A piecewise linear model (spline) was fit to the second and third phase of the PVP (log-lin, see the red spline in the right panel of Fig. 4):

– a first-order polynomial for the second phase. The intercept $\log_2 \widehat{\sigma}_0$ is in theory located at $(\tau, \log_2 \sigma_0)$, and the algebraic value of the slope is $-C$;
– the constant $\log_2 \widehat{\sigma}_\infty$ for the stationary phase.

The spline has a single knot, defined at $(\Omega, \log_2 \widehat{\sigma}_\infty)$; equations of the spline are:

$$\log_2 \sigma(t) = \log_2 \widehat{\sigma}_0 - C(t - \tau) \quad \text{if} \quad \tau \leq t \leq \Omega, \qquad (20)$$
$$= \log_2 \widehat{\sigma}_0 - C(t - \Omega) = \log_2 \widehat{\sigma}_\infty \quad \text{else.} \qquad (21)$$

The fit was computed by ordinary least squares (OLS) on the logarithm of the PVP. This minimization was performed straightforwardly with numerical optimization.

A fit is illustrated in the right panel of Fig. 4, where the empirical PVP is displayed in black and the piecewise linear fit (spline) is displayed in red. The first phase of the PVP lasts about 300 ms ($\tau = 302$ ms); from there the PVP drops quasi-linearly until $\Omega = 1.7$ s, which defines the knot of the spline at which the PVP levels off. The slope is estimated at $C = 4.6$ bit/second; the endpoint accuracy (stationary phase) is estimated at $\widehat{\sigma}_\infty = 1.6e^{-4}$ m. A coefficient of determination, computed only on the second phase, is also given ($r^2 = 0.9973$).[8]

The coefficients of determination were computed on all PVPs, see Table 1. For the PD-dataset, the quality of fits are mostly equivalent, except for the 2 conditions with ID = 2. Generally, all goodness of fits computed are high, with an average r^2 for the G-dataset of 0.95 and an average r^2 for the PD-dataset of 0.97. For the G-dataset, the quality of fit increases with the precision requirement: The mean value of r^2 increases, and its standard deviation decreases when participants are instructed to emphasize accuracy.

This can be explained by observing the left and middle panels of Fig. 4, which show that for the conditions that require low precision (Conditions #1 and #2 in the G-dataset and ID = 2 in the PD-dataset), the second phase is short and the edges of the PVPs are rounded off (likely due to the

---

[8] It can seem surprising that we give a r^2 value computed on the second phase, whereas the model was fit simultaneously on the second and third phase. But notice that by construction, the positional variance is necessarily constant after all movements have terminated. If one were to fit a linear model on the constant phase, one would invariably get a null slope and a perfect fit ($r^2 = 1$), so that the r^2 would increase mechanically with longer extension times.

Table 1 Mean and Standard deviation Summary for the PD-dataset (left) and G-dataset (right)

| W ($D \times$ ID)	Ω (μ/σ)	C (μ/σ)	r^2 (μ/σ)	Instruction	Ω (μ/σ)	C (μ/σ)	r^2 (μ/σ)
0.8 (212 × 8)	1.378/0.214	6.22/1.18	0.98/0.01	1—U-Fast	0.432/0.145	4.274/3.18	0.87/0.13
1.4 (353 × 8)	1.373/0.172	6.41/0.91	0.99/0.00	2—Fast	0.713/0.206	4.20/1.51	0.94/0.05
3.3 (212 × 6)	1.050/0.195	6.95/0.97	0.99/0.01	3—Balanced	0.851/0.249	5.43/1.61	0.96/0.05
5.5 (353 × 6)	1.077/0.158	6.93/1.18	0.99/0.01	4—Precise	0.999/0.365	5.23/0.984	0.97/0.03
14.1 (212 × 4)	0.814/0.144	6.15/1.59	0.97/0.02	5—U-Precise	1.310/0.389	5.29/0.814	0.98/0.02
23.6 (353 × 4)	0.808/0.141	6.83/1.64	0.98/0.01				
70.7 (212 × 2)	0.429/0.130	6.04/3.97	0.93/0.11				
117.8 (353 × 2)	0.527/0.143	4.82/2.76	0.95/0.03				
($\overline{\mu}/\overline{\sigma}$)	0.932/0.368	6.30/2.00	0.97/0.04		0.861/0.405	4.89/1.86	0.95/0.08
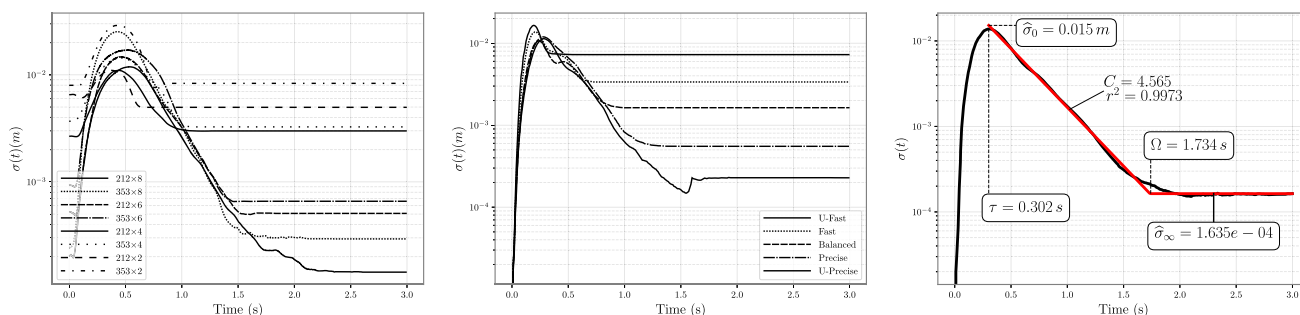
Fig. 4 Various PVPs in log-lin scale, where time and positional standard deviations are given in standard units. Left panel: PVPs for all conditions of $P8$ from the PD-dataset. Each condition should be read as D (mm) $\times$ ID. Middle panel: PVPs for all conditions of $P8$ from the G-dataset. Right panel: PVP for condition U-Precise (#5) of the G-dataset in log-lin scale with corresponding fit and estimated values of C and $\widehat{\sigma}_{\infty}$.

regularity of human produced movements). We suppose that this rounding off of the edges has more effect the shorter the second phase, leading to a degradation of the r^2.

5.2 Prediction 2: Invariance of C

Effect of D and W on C. Average and standard deviation values of C are given Table 1; values of C are between 6 and 7 bit/second except for the condition $W = 117.8$. The standard deviations are particularly high for ID $= 2$ (coefficient of variation $> 1/2$), indicating that the estimates of the slopes are unreliable in that condition.

A mixed linear regression with random intercept was computed for C with fixed effects D ($\mu = 12.7$, $\sigma = 23.5$, cv $= 1.85$), W ($\mu = 2.76$, $\sigma = 2.99$, cv $= 1.08$) and interaction term D:W ($\mu = -83.7$, $\sigma = 72.1$, cv $= 0.86$) to estimate the effect of each factor, all giving coefficients of variations (cv) close to 1, except for the fixed intercept ($\mu = 5.9$, $\sigma = 0.93$, cv $= 0.15$).

Effect of instructions on C. Average and standard deviation values of C are given Table 1. Average values of C are between 4.20 and 5.43 bit/second, the maximum being attained for the balanced condition (#3). Instructions to

emphasize accuracy lead to lower standard deviations for C: From $\sigma(C) = 3.18$ for the condition U-Fast, down to $\sigma(C) = 0.814$ for the condition U-Precise. Hence, more precise movements lead to more reliable estimates, likely due to the increase in the duration of the second phase.

To evaluate the effect of instruction on C, a one-way repeated measures ANOVA was computed ($F(1.88, 28.25) = 22$, $p_{GG} = 0.17$, $\eta_g^2 = 0.08$).

Both analyses show that these two datasets present very little to no effect of instruction or D and W on the mean value of C, thereby supporting the hypothesis that C is a constant that does not depend on the task parameters. However, reliable estimation of C seems possible only when the accuracy requirement of the task is stringent.

5.3 Gaussian trajectories

The Gaussian assumption, i.e., that observed samples at any given time follow a Gaussian distribution is central to our model. We asserted the Gaussianity of the position signal throughout the movement for both datasets by:

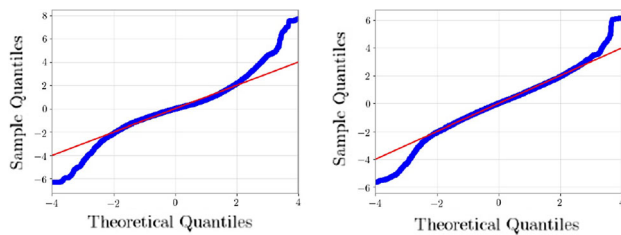1. repeating the first three steps of the PVP construction;

Fig. 5 Quantile–Quantile plots for the G-dataset (left panel) and the PD-dataset (right panel). Data considered are the reunion of the normalized data per participant and per condition computed for all sampling instants

2. starting from $t = \tau$, pooling position data into a vector for each participants for each condition and for each timestamp to create a so-called position slice;
3. normalizing each position slice.

The normalized position slices were pooled, from which a quantile–quantile plot (qqplot) was drawn with a standard normal distribution for comparison, see Fig. 5. The left panel represents the qqplot for the G-dataset, whereas the right panel represents the qqplot for the PD-dataset. Figure 5 shows the empirical distribution to be almost symmetrical, but has heavier tails than the normal distribution in both cases. For about 96% of the data, however, (in the range where points are less than 2 standard deviations away from the mean value), the empirical qqplot is an almost exact fit, see Fig. 5. The Gaussian assumption thus seems safe.

While our theoretical model of the second phase shows promising results, an equivalent model for the first phase is required to make sense of the whole movement. As explained previously, however, a useful information-theoretic model cannot lead to an increasing variance. Rather than proposing a model for the first phase that draws from a different field, we prefer at this point to present an empirical analysis of the first phase. The surprising result that we find is that the duration of the first phase is almost constant.

6 Variance increasing phase: Empirical study

D, W, ID and speed-accuracy strategy were varied, and its effect on the features of the first phase σ_0, D_τ and τ was observed. We find that movements in this first, variance increasing phase are affected primarily by D. We also find that the duration (and its variations even more) of this first phase is small in comparison with the total movement time.

6.1 Effect of D and W on τ

Values of τ are given in Table 2 and plotted in the top left panel of Fig. 6 (Table 3). Results of the regression are given in Table 4, where the estimated mean value/standard deviation,

as well as the coefficient of determination, is given for each fixed effect parameter.

6.2 Effects of D and W on D_τ

Average values of D_τ are plotted with bars in the top middle panel of Fig. 6. For comparison purposes, factor level D was also plotted on the top middle panel of Fig. 6 with a thick black line. Higher levels of D lead to larger D_τ:

– For $D = 0.212$ m, D_τ ranges from 0.126 m (ID = 2) to 0.151 m (ID = 6). The average D_τ over all ID levels is 0.141 m, (67% of D).
– For $D = 0.353$ m, D_τ ranges from 0.216 m (ID = 2) to 0.242 m (ID = 6). The average D_τ over all ID levels is 0.228 m (69% of D).

Results of the regression are given Table 4, confirming the effect of D on D_τ.

6.3 Effects of D and W on σ_0

Average values of σ_0 are plotted with bars in the top right panel of Fig. 6, which shows an increase in σ_0 with levels of D ($\overline{\sigma}_0 = 0.019$ m for $D = 0.212$ m and $\overline{\sigma}_0 = 0.029$ m for $D = 0.353$ m). This is confirmed by the regression in Table 4. The sharp "drop" in σ_0 for ID = 2, evident from Fig. 6 could indicate that aimed movement in the ID = 2 regime is different from other, more precise movements, in line with Crossman's (1983, Fig. 1) analysis for low ID movements.

Keeping only factors with a great effect size[9], the regressions of the PD-dataset (Table 4) are condensed to:

$$\tau = 0.225 + 0.343D + \varepsilon; \tag{22}$$
$$D_\tau = 0.765D + \varepsilon'; \tag{23}$$
$$\sigma_0 = 0.079D + \varepsilon''. \tag{24}$$

where ε, ε', and ε'' are error terms. This shows that τ, D_τ, and σ_0 of the first phase of the PVP depend primarily on the level of D.

6.4 Effect of instructions on τ, D_τ and σ_0

The average values of τ, D_τ, and σ_0 for the G-dataset are represented, respectively, on the left, middle, and right bottom panels of Fig. 6 and Table 5.

---

[9] As measured by a coefficient of variation much smaller than 1, where the coefficient of variation for a set of samples with observed mean μ and standard deviation σ is defined as the ratio σ/μ. There is no difficulty in choosing the cut-off rate, since coefficients of variation obtained are clear cut above or below 1. Significance testing with p-values gives identical results ($\alpha = 0.05$).
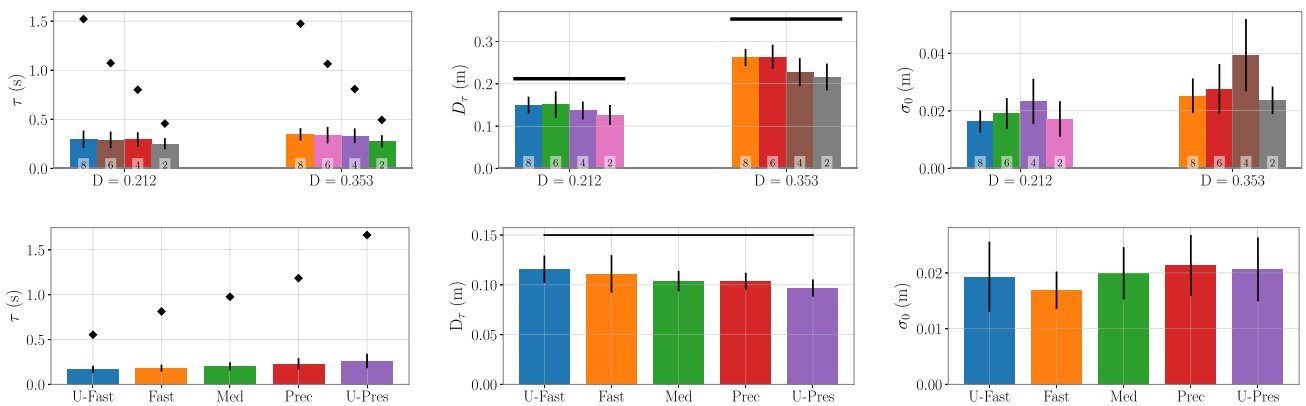
Fig. 6 Top: Effects of D, ID and W on τ (top left panel), D_τ (top middle) and σ_0 (top right) for the PD-dataset. Bars are grouped by D condition, with the 4 bars on the left corresponding to $D = 0.212$ m and the 4 bars on the right corresponding to $D = 0.353$ m. Each bar is labeled with its corresponding level of ID. Bottom: Effects of instruction on τ (bottom left panel), D_τ (bottom middle) and σ_0 (bottom right) for the G-dataset. On the τ panels, total movement time is represented with black diamonds, and on the D_τ panels, D is represented in a black thick line

Table 2 Mean (μ) and standard deviation (σ) summary for τ, D_τ and σ_0 for the PD-dataset

W ($D \times$ ID)	τ (μ/σ)	D_τ (μ/σ)	σ_0 (μ/σ)
0.8 (212 × 8)	0.296/0.089	0.150/0.020	0.016/0.004
1.4 (353 × 8)	0.347/0.063	0.262/0.020	0.025/0.006
3.3 (212 × 6)	0.291/0.085	0.151/0.031	0.019/0.005
5.5 (353 × 6)	0.340/0.084	0.264/0.028	0.028/0.009
14.1 (212 × 4)	0.295/0.074	0.137/0.021	0.023/0.008
23.6 (353 × 4)	0.333/0.075	0.227/0.033	0.039/0.006
70.7 (212 × 2)	0.252/0.057	0.126/0.024	0.017/0.006
117.8 (353 × 2)	0.276/0.062	0.216/0.032	0.024/0.005
($\overline{\mu}/\overline{\sigma}$)	0.305/0.078	0.193/0.059	0.024/0.010

The left column indicates the task conditions, where W and D are expressed in mm. All other units are standard

Table 3 Mean (μ) and standard deviation (σ) summary for τ, D_τ and σ_0 for the G-dataset

# Instruction	τ (μ/σ)	D_τ (μ/σ)	σ_0 (μ/σ)
1—U-Fast	0.167/0.041	0.115/0.014	0.019/0.006
2—Fast	0.181/0.039	0.111/0.019	0.016/0.003
3—Balanced	0.199/0.046	0.103/0.011	0.019/0.005
4—Precise	0.228/0.063	0.103/0.009	0.021/0.005
5—Precise	0.261/0.081	0.096/0.009	0.020/0.006
($\overline{\mu}/\overline{\sigma}$)	0.208/0.065	0.106/0.014	0.020/0.005

All units are standard

Table 4 Regression of τ, D_τ and σ_0

	τ (μ/σ)	D_τ (μ/σ)	σ_0 (μ/σ)
W	$-0.667/0.526$	$-0.256/0.362$	$-0.012/0.111$
D	$0.343^{***}/0.067$	$0.765^{***}/0.046$	$0.079^{***}/0.015$
$W:D$	$0.217/1.617$	$-0.319/1.114$	$-0.092/0.345$
Intercept	$0.225^{***}/0.027$	$-0.014/0.014$	$0.003/0.004$
r_m^2/r_c^2	0.14/0.79	0.78/0.82	0.28/0.33

$^* p < 0.1$; $^{**} p < 0.05$; $^{***} p < 0.01$

Table 5 One-way repeated measures ANOVA for effect of Instruction on τ, D_τ and σ_0

Observation	$F(df_{GG}, df_{GG})$	p	η_g^2	ε_{GG}
τ	$F(1.7, 25.48) = 32.4$	$p_{GG} < 10^{-6}$	0.28	0.47
D_τ	$F(2.02, 30.33) = 5.8$	$p_{GG} = 0.007$	0.22	0.51
σ_0	$F(2.44, 36.67) = 2.48$	$p_{GG} = 0.087$	0.08	0.61

– Effect on τ: Grand mean for τ is $\overline{\tau} = 0.208$ s (min $= 0.167$ s for condition U-Fast (#1), max $= 0.261$ s for condition U-Precise (#5). Results of the one-way RM ANOVA with factor Instruction are given in Table 5, showing a moderate effect ($\eta_g^2 = 0.28$) on τ. Instructions to emphasize Accuracy increase the value of τ.

– Effect on D_τ: Grand mean for D_τ is $\overline{D_\tau} = 0.106$ m (min $= 0.096$ m for condition U-Precise #5, max $= 0.115$ m for condition U-Fast #1). Results of the one-way RM ANOVA with factor Instruction are given in Table 5, showing a moderate effect ($\eta_g^2 = 0.22$) on D_τ. Instructions to emphasize speed lead to larger values of D_τ. Target distance D is represented in thick black line for comparison on the bottom middle panel of Fig. 6.

– Effect on σ_0: Grand mean for σ_0 is $\overline{\sigma_0} = 0.02$ m (min $= 0.016$ m for condition Fast #2, max $= 0.021$ m for condition Precise #4). Results of the one-way RM ANOVA with factor Instruction are given in Table 5, showing a very small effect ($\eta_g^2 = 0.08$) on σ_0. We also re-ran the ANOVA by excluding the Fast condition ($p_{GG} = 0.44$), giving η_g^2 close to null.

6.5 Discussion

Variations on τ can be considered negligible for precise aimed movements. D being the primary factor that affects τ, it is informative to compute the actual variations induced by D on τ: For $D = 0.212$ m, $\tau = 0.298$ ms, while for $D = 0.353$ m, $\tau = 0.346$ ms. The two levels of D thus induce about 50 ms variation in τ, which is less than the standard deviation associated with any one condition. In comparison, the average MT ranges from about 0.5 s to 1.5 s between conditions. Hence, at the level of change induced in MT, variations in τ are relatively negligible (5%). In other words, the variations in τ account very little in explaining the variations in observed MT. This comparison is reflected in the top left panel of Fig. 6, where variations in τ are hardly visible at the scale of MT (black diamonds).

In the case of the effect of instruction, we similarly observe that variations of τ are small compared to variations of total MT, see bottom left panel of Fig. 6, even though the effect of instruction is moderate on τ ($\eta_g^2 = 0.28$). The difference in average τ it at most about 100 ms, i.e., about 10% of the average total movement time. If one excludes the two extreme conditions, under which humans rarely perform in daily life, the difference in τ comes down to about 5% of the average total movement time, similarly to the PD-dataset. This suggests that τ can be considered constant, leading to simpler results for a comparatively small loss in modeling power in most practical cases.

The coefficients of determination computed for the mixed effect model on τ ($r_m^2 = 0.14$ and $r_c^2 = 0.79$) indicate that most of the variability of τ is actually due to differences between participants. This observation implies that if one is willing to predict τ, models linking τ to participant characteristics need to be conceived. Conversely, if one if willing to correlate participants characteristics to PVP parameters, τ should be a good candidate (e.g., older participants can be hypothesized to have, say, a larger τ).

On D_τ and the isochrony of movements. The finding that D_τ scales with D is not surprising: Assuming a constant τ, an increase in D_τ translates to an increase in average speed (D_τ / τ). The observation that average speed increases with D is in line with the so-called isochrony of movements (Guiard 2009). The findings of the G-dataset show that D_τ (and thus speed) increases with the instruction to emphasize speed (from 0.35 m/s for #5 U-precise to 0.7 m/s for #1 U-fast), in agreement with the experimental protocol.

In both datasets, the average value of D_τ had the same ratio to D, about 70% (67% and 69% for two levels of D in the PD-dataset, 71% on average for the G-dataset). Furthermore, most of the variability of D_τ is captured by the fixed effect model ($r_m^2 = 0.78$), indicating low participant differences in D_τ. It would thus be interesting to test in further work whether this proportion of about 70% continues to hold for larger variations of D, including very short and very long movements.

7 Synthesis of the two phases: Recovering Fitts' law

A combination of the two phases provides a full description of the variability of aimed movements. For consistency, it should lead to both versions of Fitts' law (1) and (2), when applied in Fitts' paradigm. We intend to show that Ω and $\widehat{\sigma}_\infty$ as defined in Sect. 5 (see also see Fig. 4) can, (respectively), be mapped to MT and σ of Fitts' law.

7.1 Synthesis of the two phases

Movement duration $\widetilde{\Omega}$ is obtained as the sum of the durations of the two phases:

$$\widetilde{\Omega} = \tau + \frac{1}{C}\log_2\left(\frac{\sigma_0}{\sigma_\infty}\right). \tag{25}$$

In Sect. 6, it was shown that $\sigma_0 \simeq \frac{0.079\,D}{\sigma_\infty}$ for the PD-dataset, giving

$$\widetilde{\Omega} = \tau + \frac{1}{C}\log_2\left(\frac{0.079\,D}{\sigma_\infty}\right), \tag{26}$$

and playing with constants, one has

$$\widetilde{\Omega} = \tau + \frac{1}{C}\log_2(0.079 \times 4.133) + \frac{1}{C}\log_2\left(\frac{D}{4.133\sigma_\infty}\right) \tag{27}$$

$$\widetilde{\Omega} = \tau' + \frac{1}{C}\log_2\left(\frac{D}{4.133\sigma_\infty}\right) \tag{28}$$

giving an expression close to (2).

To obtain the so-called nominal Fitts' law (2), we use the Gaussianity of the $\widehat{\mathbf{A}}_i$s and link the miss rate ε to the standard deviation of endpoints $\sigma_\infty = [2\sqrt{2}\,\mathrm{erf}^{-1}(1-\varepsilon)]^{-1}W$, where $\mathrm{erf}^{-1}(x)$ is the inverse Gaussian Error Function, and plug this in (25) to obtain $\breve{\Omega}$

$$\breve{\Omega} = \tau + \frac{1}{C}\log_2\left(0.079 \times 2\sqrt{2}\,\mathrm{erf}^{-1}(1-\varepsilon)\right) + \frac{1}{C}\log_2\left(\frac{D}{W}\right) \tag{29}$$

$$\breve{\Omega} = \tau''(\varepsilon) + \frac{1}{C}\log_2\left(\frac{D}{W}\right) \tag{30}$$

Miss rates ε are usually small, and the standardized methodology considers a constant 4% miss rate.[10]

The obtained equations for movement time (30) and (28) are consistent with the nominal and effective Fitts' law (1) and (2), except for the "+1" term that is missing in formulations for $\widetilde{\Omega}$ and $\breve{\Omega}$. This term has been discussed several times in the literature (Soukoreff and MacKenzie 2004; Hoffmann 2013) and is actually of little interest—the changes it induces are sensible only for very low values of the ratio D/W, where Fitts' law is known to be a poor model (Crossman and Goodeve 1983), and where the estimation of C is unreliable anyway.
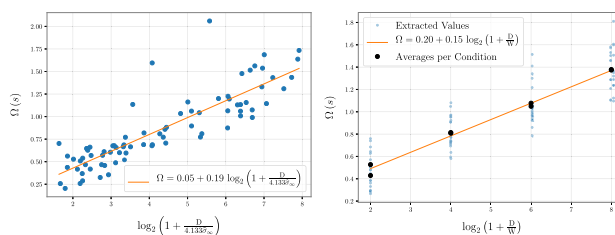
Fig. 7 Fitts' law evaluated for extracted values of Ω. Left panel: Results of Fitts' law (2) fitted on data of the G-dataset. Right panel: Results of Fitts' law (1) fitted on data of the PD-dataset. The black dots are the per-condition averages, as typically used for regression in most Fitts' law studies (see Soukoreff and MacKenzie 2004)

We evaluated Fitts' law on the set of extracted Ω and $\widehat{\sigma}_\infty$, see Fig. 7. Parameters obtained this way should be compared to those of $\widetilde{\Omega}$ and $\breve{\Omega}$, evaluated from (28) and (30) for the G-dataset as well as the PD-dataset.

PD-dataset. The overall miss rate is $\varepsilon = 2.86\%$. Using the average value of $C = 6.3$ (Table 2), one has

$$\breve{\Omega} = 0.07 + 0.16\log_2\left(\frac{D}{W}\right). \tag{31}$$

For comparison, the fitted Fitts' law using the measured Ω gives $\Omega = 0.2 + 0.15\,\mathrm{ID}$, see the left panel of Fig. 7 $(r_e^2 = 0.81,\ r^2 = 99)$.[11]

G-dataset. With the average value of $C = 4.89$, one has

$$\widetilde{\Omega} = -0.12 + 0.20\log_2\left(\frac{D}{4.133\sigma_\infty}\right). \tag{32}$$

For comparison, the fitted Fitts' law using Ω and $\widehat{\sigma}_\infty$ gives $\Omega = 0.05 + 0.19\,\mathrm{ID}$ $(r^2 = 0.74)$, see the right panel of Fig. 7.

It thus appears that, while the intercept is a little bit underestimated, the estimation of the slope is in line with empirical data. A more precise estimate for the intercept could be obtained by using a more refined model for the first phase. However, this would require many more empirical investigations, and possibly a model of some sort, which is out of the scope of this paper. A second point is that the spline does not factor in the observation that around the maximum variance point, the profiles are smooth, which further induces a bias in the intercept calculation. Furthermore, the most useful parameter in Fitts' law is the slope and not the intercept (Zhai et al. 2004).

7.2 Traditional Fitts' law evaluation

A comparison with Fitts' law parameters (34) and (33) obtained only from endpoints, with the traditional methodology confirms the consistency of our novel method.

---

[10] In fact, the standard methodology advocated in ISO (2000) uses a constant miss rate of 3.88% (see Soukoreff and MacKenzie 2004, but see also Gori et al. 2018 for a critique), in which case $2\sqrt{2}\,\mathrm{erf}^{-1}(1-\varepsilon) = 4.133$.

[11] r_e^2 and r^2 have been defined before in this paper.

PD-Dataset. Grand mean for movement time across all conditions was $\overline{\mathrm{MT}} = 0.96\,$s. Fitts' law parameters (1) were estimated via simple linear regression ($r_e^2 = 0.742$, $r^2 = 0.989$)

$$\mathrm{MT} = 0.13 + 0.17 \log_2\left(1 + \tfrac{D}{W}\right), \tag{33}$$

where r_e^2 is the coefficient of determination obtained from all MTs, and r^2 the coefficient of determination obtained after averaging MT for each condition.[12]

G-Dataset. Grand mean for movement time across all conditions was $\overline{\mathrm{MT}} = 1.04\,$s. Fitts' law parameters were obtained by simple linear regression ($r_e^2 = 0.69$)[13] on (2)

$$\mathrm{MT} = 0.28 + 0.17 \log_2\left(1 + \tfrac{D}{4.133\sigma}\right). \tag{34}$$

As in the previous case, slopes are accurately predicted and intercepts are a bit underestimated, showing the validity of our method compared to the traditional approach.

8 Conclusion

The outcome of our theoretical model, Formula (19), is a large improvement over the existing quantitative descriptions of the speed-accuracy tradeoff. It relates accuracy to speed *throughout* the movement and actually results from a joint minimization of speed and accuracy (see Theorem 1 and (11)). This is in contrast to, e.g., Fitts' law that can only describe endpoints, or optimal control models for goal-directed, which often require the time horizon to be preselected (Todorov 1998) (but see Harris and Wolpert 1998; Guigon et al. 2008; Berret and Jean 2016; Tanaka et al. 2006) and which function with various cost functions (Todorov 1998; Ning Qian et al. 2013; Flash and Hogan 1985) that can be considered ad hoc. It also has a direct operational implication: The variance decreases exponentially at a constant rate C during the second phase, and this is what largely predicts the movement time.

As discussed by Tanaka et al. (2006), two movement expressed as the solutions of an optimization problem where (1) the duration of movement is fixed and the variability is minimized, or (2) the variability is fixed and the duration of

movement is minimized are not *a priori* equivalent.[14] The minimization of mutual information that we propose in this work, jointly optimizing for speed and accuracy, actually makes that equivalence.

The functions **f** and **g** were not defined beforehand, but left unspecified, to be determined as part of the optimization problem. It turns out that the resulting optimal scheme is linear, surprisingly simple and does not require memory beyond the obvious fact that the limb maintains its position in the steady state. This finding provides support for the many linear schemes found in the motor control literature.

As a final observation, let us recall that originally Fitts' law was conceived via a vague analogy with Shannon's capacity formula (1948), since then deemed flawed (see, e.g., Gori et al. 2018 for a brief history). The information-theoretic framework appears naturally here to solve the aiming problem, upholding that information theory can be a useful tool to model human performance (Chan and Childress 1990; Sheridan and Ferrell 1974). In addition, the relevant signal never actually needs to be transformed into the vague notion of ID (Guiard and Olafsdottir 2011), measured in bits; mutual information is here expressed as a tangible ratio of two spatial variances.

A final advantage comes from the method of PVPs described here. Segmenting movements correctly is notoriously hard (Teasdale et al. 1993). Specifically, determining when a movement (or submovement) ends is particularly difficult, since movements performed with high levels of accuracy usually end very smoothly.[15] The method of PVPs dispenses the experimenter of the complicated task of precisely determining stopping times. It also handles trajectories indifferently of the associated control, be it discrete continuous or intermittent.

This work provides a new theoretical model for voluntary movements, based on the study of positional variance profiles (PVPs), which takes into account: (a) The variability of human produced movements; (b) a feedback mechanism, essential for reliable aiming; and (c) intermittent control that becomes continuous at the limit. Empirical evidence shows that multi-joint goal-directed movements lead to unimodal PVPs: A first variance-increasing phase of approximately constant duration is followed by a variance-decreasing second phase that lasts until an appropriate accuracy level is reached. We established that:

---

[12] The r^2 is included for comparison purposes with figures from the literature—most Fitts' law studies average data prior to regression and usually only r^2's are given.

[13] The averaged coefficient of determination r^2 cannot be computed here as there is a different level of ID_e for each block, even between blocks performed in the same condition.

---

[14] Interestingly, these two partial optimizations correspond to the two different empirical paradigms of Schmidt and Fitts, see, e.g., Plamondon and Alimi (1997).

[15] In fact, a movement never truly ends, as keeping the position stationary over time requires a control of some sort.

1. The problem of aiming, during the second component, can be reduced to that of transmitting information from a source (position at the end of the first component) to a destination (limb extremity, cursor) over a channel perturbed by Gaussian noise with the presence of a noiseless feedback link.
2. Using an optimal scheme, in the sense that the transmitted information from source to destination is maximized at each step, we showed that positional variance can decrease at best exponentially during the second phase, as summarized by (19), at a constant rate C.

Beyond the theoretical implications of this work, we believe that our model can have a direct impact on the current practices in the evaluation of pointing devices in HCI. Indeed, this work suggests an alternative to Fitts' law for the evaluation of pointing devices, by:

1. instructing users to point exactly on a line as fast as they can (i.e., the case when C is the most reliably estimated),
2. computing PVPs and estimating C.

C being constant across speed-accuracy instructions and levels of D and W, there is no need for repetitive measurements over several D and W conditions. This new method would have several practical advantages over a traditional Fitts' law style evaluation.

– the model provides a description throughout most of the trajectory, whereas Fitts' law can only inform about endpoints, allowing finer analyses (e.g., rather than asking if one group is faster than the other, one can for investigate whether some group composed of a special population displays a reduction in variance that is subexponential);
– the evaluation of pointing performance with PVPs does not require a predefined width as in Fitts' task; allowing evaluation in new cases (e.g., for users with strong motor impairment, a predefined width is unfeasible Davies et al. 2014);
– there is no need to determine the (difficultly tractable) stoppage time of a movement; this point may be particularly useful with motor-impaired populations;
– since the slope can be estimated from a single experimental condition, using PVPs to estimate Fitts' law parameters saves the experimenter two factors (D and W), saving valuable time. More experimental and methodological work needs to be conducted, however, to understand how variability in the estimation of C can be reduced.

Further practical work should determine whether this method is indeed applicable, particularly whether the accuracy of the estimation of C can be further enhanced.

It would also be interesting to observe the effect on the kinematics if one modified some part of the scheme. For example, what would happen were the feedback nonideal? Can this serve to model the movement of some motor impaired patients, e.g., for pathologies which have afflicted the peripheral nervous system? Similarly, what would happen if there were no scaling α_i's or if they were poorly "chosen"? Could this serve to model tremors (e.g., with α_i's much too large)?

A Appendix

A.1 Segmentation algorithm

The algorithm works in the following steps (using the preprocessed time series as presented in Sect. 4):

1. Identify time instants $\{t_{0,i}\}_{i=1}^n$ when position crosses half of the distance between start and target while maintaining a positive speed, thereby identifying n movements[16];
2. Compute the velocity profile (from the position profile), normalize it with respect to its maximum value, to determine the start of each movement via thresholding (go back in time from $t_{0,i}$ until the normalized velocity reaches, say, 1%);
3. Look for "dwell periods" after $t_{0,i}$, i.e., intervals where the absolute value of the normalized velocity is below 1% and when the current position is above the one obtained at $t_{0,i}$. The latest instant of the last dwell period is the end of the movement.[17]

A.2 τ Box-plot for each participant

The Box plot for τ for each participant is displayed Fig. 8.

---

[16] In the case of the reciprocal paradigm, this amounts to removing all movements going right to left and keeping only movements going left to right. The other half of movements can be retrieved by inverting the trajectory and performing the exact same operations. In this paper, we only keep movements going left to right, to eliminate potential differences between left to right and right to left movements which we are not concerned with.

[17] Right after this instant, either a new movement begins (in a reciprocal task), or the cursor is moved back to the start position (in a discrete task).
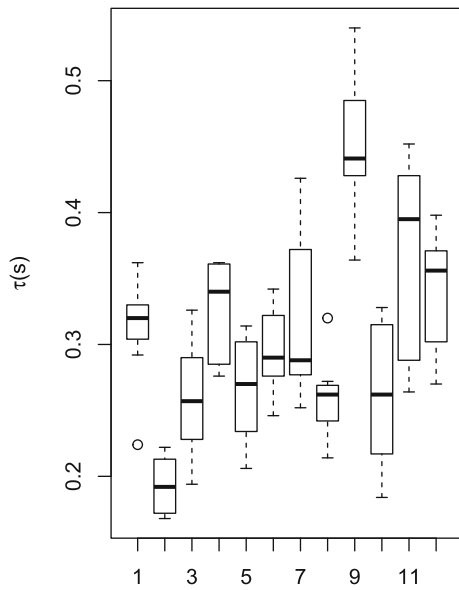
Fig. 8 τ (s) plotted for each participant of the PD-dataset

A.3 Mathematical proofs

In the proofs, we use the natural logarithm (log) rather than its base-2 form ($\log_2$), for simplicity and consistency with information-theoretic textbooks and is equivalent barring the change of units from bits ($\log_2$) to nats (log) (Cover and Thomas 2012).

Proof of Theorem 1 We use well-known ingredients from information-theory (Cover and Thomas 2012). For inequality (10a):

$$I(\mathbf{A}; \widehat{\mathbf{A}}_n) = H(\mathbf{A}) - H(\mathbf{A}|\widehat{\mathbf{A}}_n) \tag{35}$$

$$= H(\mathbf{A}) - H(\mathbf{A} - \widehat{\mathbf{A}}_n|\widehat{\mathbf{A}}_n) \tag{36}$$

$$\geq H(\mathbf{A}) - H(\mathbf{A} - \widehat{\mathbf{A}}_n) \tag{37}$$

$$\geq H(\mathbf{A}) - \frac{1}{2}\log\left(2\pi e \mathbb{E}[(\mathbf{A} - \widehat{\mathbf{A}}_n)^2]\right) \tag{38}$$

$$= \frac{1}{2}\log\frac{\sigma_0^2}{D_n}, \tag{39}$$

where (35) is by definition of mutual information; (36) because of the conditioning by $\widehat{\mathbf{A}}_n$; (37) because conditioning reduces entropy; (38) because the Gaussian distribution maximizes entropy under power constraints and by the entropy formula for a Gaussian distribution; (39) by definition of the distortion and the entropy formula for a Gaussian distribution.
For inequality (10b):

$$I(\mathbf{A}; \widehat{\mathbf{A}}_n) \leq I(\mathbf{A}; \mathbf{Y}^n) \tag{40}$$

$$= H(\mathbf{Y}^n) - H(\mathbf{Y}^n|\mathbf{A}) \tag{41}$$

$$= \sum_{i=1}^{n}\left[H(\mathbf{Y}_i|\mathbf{Y}^{i-1}) - H(\mathbf{Y}_i|\mathbf{Y}^{i-1}, \mathbf{A})\right] \tag{42}$$

$$= \sum\left[H(\mathbf{Y}_i|\mathbf{Y}^{i-1}) - H(\mathbf{Y}_i|\mathbf{X}_i)\right] \tag{43}$$

$$\leq \sum\left[H(\mathbf{Y}_i) - H(\mathbf{Z}_i)\right] \tag{44}$$

$$\leq \sum\left[\frac{1}{2}\log(2\pi e(P_i + N)) - \frac{1}{2}\log(2\pi eN)\right] \tag{45}$$

$$\leq \sum_{i=1}^{n}\left[\frac{1}{2}\log(1 + P_i/N)\right] \leq nC \tag{46}$$

where (40) is by the data processing inequality (Cover and Thomas 2012) applied the Markov chain $\mathbf{A} \longrightarrow \mathbf{Y}^i \longrightarrow \mathbf{g}(\mathbf{Y}^i) = \widehat{\mathbf{A}}^i$; (41) by definition; (42) by applying the chain rule (Cover and Thomas 2012) to both terms; (43) by design of the feedback scheme; (44) because conditioning reduces entropy for the first term a,d by virtue of the AWGN model for the second term; (45) because the Gaussian distribution maximizes entropy and $\mathbf{X}_i$ and $\mathbf{Z}_i$ are independent; (46) by the concavity of the logarithm function. $\square$

Proof of Lemma 1 The proof consists of finding the conditions that make the inequalities in the proof of Theorem 1 equalities. Equality in (37) is equivalent to condition (4); Equality in (38) is equivalent to $\mathbf{A} - \widehat{\mathbf{A}}^i$ Gaussian; Equality in (40) is equivalent to $H(\mathbf{A}|\mathbf{Y}^i) = H(\mathbf{A}|\mathbf{Y}^i, \mathbf{g}(\mathbf{Y}^i)) = H(\mathbf{A}|\mathbf{g}(\mathbf{Y}^i))$, so that $\mathbf{Y}^i \longrightarrow \mathbf{g}(\mathbf{Y}^i) \longrightarrow \mathbf{A}$ form a Markov chain (Cover and Thomas 2012), leading to condition (5); Equality in (44) is equivalent to condition (3); Equality in (45) means the Y_i's are Gaussian; Equality in (46) leads to condition (2) by concavity of the logarithm. Finally, $\mathbf{X}_i$ is Gaussian as the result of the sum of two Gaussians $\mathbf{Y}_i$ and $\mathbf{Z}_i$, and so is $\widehat{\mathbf{A}}_i$ as the sum of $\mathbf{A}$ and $\mathbf{A} - \widehat{\mathbf{A}}_i$. This finally yields condition (1). $\square$

Proof of Theorem 3 We start by considering $\mathbf{X}_i = \mathbf{f}(\mathbf{Y}^{i-1}, \mathbf{A})$, which should be independent of $\mathbf{Y}_{i-1}, \forall i$ by condition (4) of Lemma 1. This implies the decorrelation

$$\mathbb{E}[\mathbf{f}(\mathbf{Y}^{i-1}, \mathbf{A})(\mathbf{Y}_{i-1})] = 0, \quad \forall i. \tag{47}$$

Since $\mathbf{X}_i$ is a function of two Gaussians $\mathbf{A}$ and $\mathbf{Y}^{i-1}$, the conditional expectation $\mathbf{X}_i = \mathbb{E}[\mathbf{X}_i|\mathbf{A}, \mathbf{Y}^{i-1}]$ is linear, hence $\mathbf{X}_i = \alpha_i(\mathbf{A} - \tilde{\mathbf{f}}(\mathbf{Y}^{i-1}))$. Plugging this in (47) makes for a direct application of the orthogonality principle, showing that $\tilde{\mathbf{f}} = \mathbb{E}[\mathbf{A}|\mathbf{Y}^{i-1}] = \mathbf{g}(\mathbf{Y}^{i-1})$. $\square$

Proof of Theorem 4 The goal of the proof is to evaluate $\mathbb{E}[\mathbf{A}|\mathbf{Y}^{i-1}]$. We first use the operational formula from the orthogonality principle $\mathbb{E}[\mathbf{A}|\mathbf{Y}^{i-1}] = \mathbb{E}[\mathbf{A}(\mathbf{Y}^{i-1})^t] \mathbb{E}[\mathbf{Y}^{i-1}(\mathbf{Y}^{i-1})^t]^{-1}\mathbf{Y}^{i-1}$. Because the channel outputs are independent [conditions (3) and (5) from Lemma 1], and input powers are identical (conditions 2 from Lemma 1), $\mathbb{E}[\mathbf{Y}^{i-1}(\mathbf{Y}^{i-1})^t]^{-1} = (P + N)^{-1}\mathbb{I}$, where $\mathbb{I}$ is the identity

matrix of size $i-1$. Then, let $\mathbf{A}_i = \mathbf{X}_i/\alpha_i$ be the unscaled version of $\mathbf{X}_i$ and notice that $\mathbf{A} - \mathbf{A}_i = \mathbf{g}(\mathbf{Y}^{i-1})$. As the channel outputs are independent, we immediately have that $\mathbb{E}[(\mathbf{A} - \mathbf{A}_i)Y_i] = 0$, hence $\mathbb{E}[\mathbf{A}Y_i] = \mathbb{E}[\mathbf{A}_iY_i]$.

Combining both results, we get

$$\mathbb{E}[\mathbf{A}|\mathbf{Y}^{i-1}] = (P+N)^{-1}\mathbb{E}[\mathbf{A}(\mathbf{Y}^{i-1})^t]\mathbb{I}\mathbf{Y}^{i-1} \tag{48}$$

$$= (P+N)^{-1}\sum_{j=1}^{i-1}\mathbb{E}[\mathbf{A}\mathbf{Y}_j]\mathbf{Y}_j \tag{49}$$

$$= (P+N)^{-1}\sum_{j=1}^{i-1}\mathbb{E}[\mathbf{A}_j\mathbf{Y}_j]\mathbf{Y}_j \tag{50}$$

$$= \sum_{j=1}^{i-1}\mathbb{E}[\mathbf{A}_j|\mathbf{Y}_j] \tag{51}$$

where

$$\mathbb{E}[\mathbf{A}_i|\mathbf{Y}_i] = (P+N)^{-1}\mathbb{E}[\mathbf{A}_i\mathbf{Y}_i]\mathbf{Y}_i \tag{52}$$

$$= (P+N)^{-1}\mathbb{E}[\mathbf{X}_i/\alpha_i \cdot (\mathbf{X}_i+\mathbf{Z}_i)]\mathbf{Y}_i \tag{53}$$

$$= \frac{1}{\alpha_i}\frac{P}{P+N}\mathbf{Y}_i \tag{54}$$

$$\square$$

Proof of Theorem 5 First, notice that we can write D_i as

$$D_i = \mathbb{E}[(\mathbf{A} - \mathbb{E}[\mathbf{A}|\mathbf{Y}^i])^2] \tag{55}$$

$$= \mathbb{E}[(\mathbf{A} - (\mathbb{E}[\mathbf{A}|\mathbf{Y}^{i-1}] + \mathbb{E}[\mathbf{A}|\mathbf{Y}_i]))^2] \tag{56}$$

$$= D_{i-1} - \mathbb{E}[(\mathbb{E}[\mathbf{A}|\mathbf{Y}_i])^2] \tag{57}$$

Using (54), one has

$$D_i = D_{i-1} - \frac{1}{\alpha_i^2}\frac{P^2}{P+N}.$$

Finally, notice that $D_{i-1} = \mathbb{E}[(\mathbf{A} - \mathbb{E}[\mathbf{A}|\mathbf{Y}^{i-1}])^2] = \mathbb{E}[(\mathbf{A}_i)^2] = P/\alpha_i^2$, to see that

$$D_i = D_{i-1}\left(1 - \frac{P}{P+N}\right) = \frac{D_{i-1}}{1+P/N}$$

The closed form for the distortion is obtained by applying this equation recursively, starting from $D_0 = \mathbb{E}[\mathbf{A}^2] = \sigma_0^2$:

$$D_i = \frac{\sigma_0^2}{(1+P/N)^i}. \tag{58}$$

Finally, we evaluate α_i (with $\alpha_0 = \frac{\sqrt{P}}{\sigma_0}$):

$$\alpha_i = \sqrt{\frac{P}{D_i}} = \frac{\sqrt{P}}{\sigma_0}(1+P/N)^{i/2} = \alpha_0(1+P/N)^{i/2}. \tag{59}$$

$$\square$$

References

Berret B, Jean F (2016) Why don't we move slower? The value of time in the neural control of action. J Neurosci 36(4):1056–1070

Bizzi E, Accornero N, Chapple W, Hogan N (1984) Posture control and trajectory formation during arm movement. J Neurosci 4(11):2738–2744

Buchanan TS, Lloyd DG, Manal K, Besier TF (2004) Neuromusculoskeletal modeling: estimation of muscle forces and joint moments and movements from measurements of neural command. J Appl Biomech 20(4):367–395

Bullock D, Grossberg S (1988) Neural dynamics of planned arm movements: emergent invariants and speed-accuracy properties during trajectory formation. Psychol Rev 95(1):49

Card SK, English WK, Burr BJ (1978) Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. Ergonomics 21(8):601–613

Carlton LG (1992) Visual processing time and the control of movement. Adv Psychol 85:3–31

Chan R, Childress D (1990) On a unifying noise-velocity relationship and information transmission in human-machine systems. IEEE Trans Syst Man Cybern 20(5):1125–1135

Chua R, Elliott D (1993) Visual regulation of manual aiming. Hum Mov Sci 12(4):365–401

Cover T, Thomas J (2012) Elements of information theory. Wiley, New York

Crossman ERFW (1957) The speed and accuracy of simple hand movements. The nature and acquisition of industrial skills

Crossman ERFW, Goodeve PJ (1983) Feedback control of hand-movement and Fitts' law. Q J Exp Psychol 35(2):251–278 First presented at the Oxford Meeting of the Experimental Psychology Society, July 1963First presented at the Oxford Meeting of the Experimental Psychology Society, July 1963

Darling WG, Cooke JD (1987) Changes in the variability of movement trajectories with practice. J Mot Behav 19(3):291–309

Davies TC, AlManji A, Stott NS (2014) A cross-sectional study examining computer task completion by adolescents with cerebral palsy across the manual ability classification system levels. Dev Med Child Neurol 56(12):1180–1186

Desmurget M, Grafton S (2000) Forward modeling allows feedback control for fast reaching movements. Trends Cogn Sci 4(11):423–431

Elia N (2004) When bode meets Shannon: control-oriented feedback communication schemes. IEEE Trans Autom Control 49(9):1477–1488

Elias P (1957) Channel capacity without coding. Proc Inst Radio Eng 45:381–381

Elliott D, Garson RG, Goodman D, Chua R (1991) Discrete vs. continuous visual control of manual aiming. Hum Mov Sci 10(4):393–418

Elliott D, Chua R, Pollock BJ, Lyons J (1995) Optimizing the use of vision in manual aiming: the role of practice. Q J Exp Psychol Sect A 48(1):72–83

Elliott D, Helsen WF, Chua R (2001) A century later: Woodworth's (1899) two-component model of goal-directed aiming. Psychol Bull 127(3):342

Elliott D, Hansen S, Grierson LEM, Lyons J, Bennett SJ, Hayes SJ (2010) Goal-directed aiming: two components but multiple processes. Psychol Bull 136(6):1023

Elliott D, Lyons J, Hayes SJ, Burkitt JJ, Roberts JW, Grierson LEM, Hansen S, Bennett SJ (2017) The multiple process model of goal-directed reaching revisited. Neurosci Biobehav Rev 72:95–110

Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 47(6):381

Fitts PM, Peterson JR (1964) Information capacity of discrete motor responses. J Exp Psychol 67(2):103

Flanagan JR, Rao AK (1995) Trajectory adaptation to a nonlinear visuomotor transformation: evidence of motion planning in visually perceived space. J Neurophysiol 74(5):2174–2178

Flash T, Hogan N (1985) The coordination of arm movements: an experimentally confirmed mathematical model. J Neurosci 5(7):1688–1703

Franklin DW, Wolpert DM (2011) Computational mechanisms of sensorimotor control. Neuron 72(3):425–442

Gallager RG, Nakiboglu B (2010) Variations on a theme by Schalkwijk and Kailath. IEEE Trans Inf Theory 56(1):6–17

Gawthrop P, Loram I, Lakie M, Gollee H (2011) Intermittent control: a computational theory of human control. Biol Cybern 104(1):31–51

Gori J, Rioul O, Guiard Y (2018) Speed-accuracy tradeoff: a formal information-theoretic transmission scheme (FITTS). ACM Trans Comput Hum Interact 25(5):27:1–27:33

Guiard Y (2009) The problem of consistency in the design of Fitts' law experiments: consider either target distance and width or movement form and scale. In: Proceedings SIGCHI conference, pp 1809–1818

Guiard Y, Olafsdottir HB (2011) On the measurement of movement difficulty in the standard approach to Fitts' law. PLoS ONE 6(10):e24389

Guiard Y, Olafsdottir HB, Perrault ST (2011) Fitt's law as an explicit time/error trade-off. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 1619–1628

Guigon E, Baraduc P, Desmurget M (2007) Computational motor control: redundancy and invariance. J Neurophysiol

Guigon E, Baraduc P, Desmurget M (2008) Optimality, stochasticity, and variability in motor behavior. J Comput Neurosci 24(1):57–68

Gutman SR, Gottlieb GL (1992) Basic functions of variability of simple pre-planned movements. Biol Cybern 68(1):63–73

Gutman SR, Latash ML, Almeida GL, Gottlieb GL (1993) Kinematic description of variability of fast movements: analytical and experimental approaches. Biol Cybern 69(5–6):485–492

Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. Nature 394(6695):780

Hoffmann ER (2013) Which version/variation of Fitts' law? A critique of information-theory models. J Mot Behav 45(3):205–215

Hogan N (1985) The mechanics of multi-joint posture and movement control. Biol Cybern 52(5):315–331

Kailath T (1980) Linear systems, vol 156. Prentice-Hall, Englewood Cliffs

ISO 9241-9:2000. Ergonomic requirements for office work with visual display terminals (VDTs)—Part 9: requirements for non-keyboard input devices. Standard, International Organization for Standardization, Geneva, Switzerland, February 2000

Keele SW (1968) Movement control in skilled motor performance. Psychol Bull 70(6p1):387

Ketcham C, Stelmach G (2004) Movement control in the older adult. In: Technology for adaptive aging. National Academies Press (US)

Khan MA, Franks IM, Elliott D, Lawrence GP, Chua R, Bernier P-M, Hansen S, Weeks DJ (2006) Inferring online and offline processing of visual feedback in target-directed movements from kinematic data. Neurosci Biobehav Rev 30(8):1106–1121

Lai S-C, Mayer-Kress G, Sosnoff JJ, Newell KM (2005) Information entropy analysis of discrete aiming movements. Acta Psychol 119(3):283–304

Mehta B, Schaal S (2002) Forward models in visuomotor control. J Neurophysiol 88(2):942–953

Meyer D, Abrams R, Kornblum S, Wright C, Keith S (1988) Optimality in human motor performance: ideal control of rapid aimed movements. Psychol Rev 95(3):340

Müller J, Oulasvirta A, Murray-Smith R (2017) Control theoretic models of pointing. ACM Trans Comput Hum Interact (TOCHI) 24(4):27

Pélisson D, Prablanc C, Goodale MA, Jeannerod M (1986) Visual control of reaching movements without vision of the limb. Exp Brain Res 62(2):303–311

Plamondon R, Alimi AM (1997) Speed/accuracy trade-offs in target-directed movements. Behav Brain Sci 20(2):279–303

Polit A, Bizzi E (1978) Processes controlling arm movements in monkeys. Science 201(4362):1235–1237

Polit A, Bizzi E (1979) Characteristics of motor programs underlying arm movements in monkeys. J Neurophysiol 42(1):183–194

Proteau L, Marteniuk RG, Girouard Y, Dugas C (1987) On the type of information used to control and learn an aiming movement after moderate and extensive training. Hum Mov Sci 6(2):181–199

Ning Qian Yu, Jiang Z-PJ, Mazzoni P (2013) Movement duration, Fitts's law, and an infinite-horizon optimal feedback control model for biological motor systems. Neural Comput 25(3):697–724

Rosenbaum DA (2009) Human motor control. Academic Press, London

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:623–656

Sheridan TB, Ferrell WR (1974) Man-machine systems; Information, control, and decision models of human performance. MIT Press, Cambridge

Simmons G, Demiris Y (2005) Optimal robot arm control using the minimum variance model. J Robotic Syst 22(11):677–690

Soukoreff R, MacKenzie I (2004) Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. Int J Hum Comput Stud 61(6):751–789

Tanaka H, Krakauer JW, Qian N (2006) An optimization principle for determining movement duration. J Neurophysiol

Tatikonda SC (2000) Control under communication constraints. PhD thesis, Massachusetts Institute of Technology

Teasdale N, Bard C, Fleury M, Young DE, Proteau L (1993) Determining movement onsets from temporal series. J Mot Behav 25(2):97–106

Todorov E, Jordan M (2002) Optimal feedback control as a theory of motor coordination. Nat Neurosci 5(11):1226–1235

Todorov EV (1998) Studies of goal directed movements. PhD thesis, Massachusetts Institute of Technology

Van der Meulen JHP, Gooskens RHJM, Denier Van der Gon JJ, Gielen CCAM, Wilhelm K (1990) Mechanisms underlying accuracy in fast goal-directed arm movements in man. J Mot Behav 22(1):67–84

Welford AT (1960) The measurement of sensory-motor performance: survey and reappraisal of twelve years' progress. Ergonomics 3(3):189–230

Wiener N (1961) Cybernetics or control and communication in the animal and the machine. MIT Press, Cambridge. 1948 2nd revised ed

Wolpert DM (1997) Computational approaches to motor control. Trends Cogn Sci 1(6):209–216

Woodworth RS (1899) Accuracy of voluntary movement. Psychol Rev Monogr Suppl 3(3):i

Zelaznik HN, Hawkins B, Kisselburgh L (1983) Rapid visual feedback processing in single-aiming movements. J Mot Behav 15(3):217–236

Zhai S (2004) Characterizing computer input with Fitts' law parameters–the information and non-information aspects of pointing. Int J Human Comput Stud 61(6):791–809

Zhai S, Kong J, Ren X (2004) Speed-accuracy tradeoff in Fitts' law tasks–on the equivalency of actual and nominal pointing precision. Int J Hum Comput Stud 61(6):823–856

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.