

# Pointing models for users operating under different speed accuracy strategies

Julien Gori

July 8, 2024

## 1 Introduction

Fitts' law predicts the time it takes one to select a target with an input device *e.g.*, with a desktop mouse [?]. Classically, pointing time, also called movement time MT is modeled with Fitts' law [?, ?, ?]

$$MT = a + b \log_2(1 + D/W), \quad (1)$$

$$= a + b ID \quad (2)$$

where  $D$  and  $W$  are geometric parameters (size and distance of the target),  $ID$  is the index of difficulty, and  $a$  and  $b$  are the estimated parameters of the linear model. Fitts' law plays an important role in HCI because pointing is a ubiquitous interaction, namely when evaluating input techniques [?] and when designing interfaces [?].

Movement modeling has advanced greatly since Fitts' seminal work [?] and the introduction of Fitts' law in HCI [?]. While Fitts' law only describes *endpoints*, motor control models now propose description of entire *trajectories* produced during pointing tasks [?, ?, ?, ?, ?]. While Fitts' law only describes *means* of movement times, several models now describe *distribution* of movement times using asymmetric distributions [?, ?, ?, ?, ?, ?].

In comparison, models of pointing when users operate under different speed accuracy strategies have rarely been investigated. Such models are useful when one needs to describe how users point with various incentives *i.e.*, favoring accuracy or speed, for example in contexts where safety matters critically [?]. One notable exception is the Weighted Homographic (WWho) model of Guiard and Rioul [?]. The WWho model was recently used as a low level control module in hierarchical models [?, ?], where a high level controller decides on the pointing strategy — the controller learns to be more cautious in contexts where the cost to recover from errors is higher.

However, the WHo model is not the appropriate tool to describe average user performance, since it is a model of *ultimate* performance: it does not describe the entire distribution of data, but instead, it describes the best average precision and speed one can obtain. In Guiard and Rioul’s work, the WHo model is fitted only on the best 10 mean block measures, out of 400! Hence, works that use the WHo model use an overly optimistic model of movement of mean movement time instead of a realistic model that describes distributions of movement time.

The goal of this work is to suggest and compare simple candidate models for distribution of movement times that incorporate user strategies. We first describe how the index of difficulty and movement time are *jointly* distributed for various user strategies. As expected, both index of difficulty and movement time increase when the strategy favors accuracy, and correlation between both measures is substantial. We then present three candidate models that generate pairs of correlated indexes of difficulty and movement time for a target user accuracy. We find out that our model based on Gaussian copulas has the best performance.

**Notations**  $\mathcal{N}(\mu, \Sigma)$  refers to the (multivariate) Gaussian distribution with mean  $\mu$  and covariance matrix  $\Sigma$ ,  $\mathcal{E}(\lambda)$  refers to the exponential distribution with mean  $\lambda$ ,  $\mathcal{U}[a, b]$  refers to the uniform distribution on  $[a, b]$ .

## 2 Background

- le gaussian assumption est pas forcément populaire. Trouver des papiers qui l’utilisent? Le seul truc qu’on puisse dire c’est qu’OLS est utilisé est OLS est optimal que quand les résidus sont gaussiens. Aussi le t-test pour la régression suppose soit large sample size, soit normalité des résidus = normalité des distributions temporelles.
- Voir pour utiliser WHo pour numériser la stratégie

### 2.1 Participant strategies and the effective index of difficulty

Fitts’ law experiments are usually conducted to investigate/evaluate the speed accuracy tradeoff. To do so, the experimenter manipulates accuracy by modifying  $W$  and observes speed of the response by measuring  $MT$  in Equation 1. However, it is well known that this manipulation is imperfect: users underutilize *easy* (low ID) targets and over-utilize *hard* (high

ID) targets [?, ?]. A solution, due to Crossman and popularized in HCI by Mackenzie [?] is to compute an effective index of difficulty  $ID_e$  as a *post-hoc* adjustment of difficulty as a function of the effectively measured standard deviation of endpoints  $\sigma$

$$MT = a + b ID_e \quad (3)$$

$$= a + b \log_2 \left( 1 + \frac{D}{4.133\sigma} \right). \quad (4)$$

Several researchers, beginning with Mackenzie [?] have investigated how people perform Fitts’ task when instructed to emphasize either speed or accuracy. As expected, participants are able to modify their strategies, as reflected by changes in  $ID_e$  for the same nominal ID level. For example in Mackenzie’s experiment, on average participants erred up to 20% in the speed condition, but erred close to 0% in the accuracy condition. Guiard [?] later performed a similar experiment, but without explicit width (participants had to aim towards a line) with 5 instructions (Ultra Fast, Fast, Balanced, Accurate, Ultra Accurate). Other similar experiments include [Ufuk, Wolfgang, Shota].

## 2.2 Classical modeling: the Gaussian assumption

Fitts’ law is classically understood to be a law of averages of minimum movement times [?], fitted with linear regression / ordinary least squares. The implicit assumption is that movement time is Gaussian distributed [?]:

$$MT \sim \mathcal{N}(\beta \mathbf{x}, s^2), \quad (5)$$

$$\beta = [a, b], \quad (6)$$

$$\mathbf{x} = [1, ID]^t, \quad (7)$$

$$(8)$$

where  $a$  and  $b$  are the parameters of Fitts’ law Equation 1, and  $s$  is the standard deviation of MT. This model, while popular, is known to be poor in particular when the experiment is not a controlled one, as shown by the “Fitts in the wild” study by Chapuis *et al.* [?], but even in controlled experiments there is ample evidence that movement time distributions are asymmetric, with long right tails [?, ?, ?].

## 2.3 The Exponentially Modified Gaussian noise model

It has been shown that the marginal distribution of MTs for a given ID/ $ID_e$  level is distributed according to an asymmetric, positive (right) skewed [?,

?, ?, ?, ?, ?, ?] distribution where the standard deviation increases with the mean value [?, ?, ?]. While several distributions may be possible, including Gamma, Gumbel and Lognormal distributions, we follow Gori and colleagues [?, ?, ?], and Zhao *et al.* [?] who considered an Exponentially Modified Gaussian (EMG) noise model as MT distribution: it satisfies all the mentioned properties, has parameters that are easy to interpret, has shown good performance (including when modeling steering data [?]), and there exist tools to aid in parameter estimation [?, ?].

The EMG model can be interpreted as an extension of the Gaussian model Equation 5, where an *exponential* random variable  $E$  is added to the *Gaussian*  $G$ :

$$MT = G + E, \quad (9)$$

$$G \sim \mathcal{N}(\beta \mathbf{x}, s), \quad (10)$$

$$E \sim \mathcal{E}(\lambda \mathbf{x}), \quad (11)$$

where  $\lambda = [\lambda_0, \lambda_1]$  represents the asymmetric component of noise. Depending on the parameter values, the EMG distribution can behave like a Gaussian ( $s \gg \lambda \mathbf{x}$ ), like an exponential ( $s \ll \lambda \mathbf{x}$ ) or in between [?].

## 2.4 The WHo model

The Who model by Guiard and Rioul [?] is an axiomatic model that conforms to the following 5 axioms:

- a1 There is a minimum movement time  $MT_0$  achievable
- a2 If we define relative precision as  $y = \sigma/D$  with  $\sigma$  the standard deviation of endpoints, then there is a minimum relative precision  $y_0$  achievable.
- a3 The function  $f$  that links MT to  $y$ ,  $y = f(MT)$  is decreasing and convex
- a4 For a given participant effort  $k$ , there is an operation such that  $MT \odot y = k$
- a5 Effort invested is never total

The WHo model

$$(y - y_0)^{1-\alpha}(MT - MT_0)^\alpha = k \quad (12)$$

satisfies a(1) – a(4), and is fit on the convex hull of the scatterplot in the (MT,  $y$ ) space to satisfy a(5). The WHo model is *not* a description of distribution of movement times, instead it characterizes the best aggregate samples (axioms 2 and 5). Guiard and Rioul never described what the distribution of movement times could look like within their WHo model.

## 2.5 Gaussian copulas

(Bivariate) Copulas are a statistical tool that describe the dependence between (two) random variables. Concretely, a multivariate joint distribution (*e.g.*, weight and height of humans) can be written in terms of univariate marginal distribution functions (width, height) and a copula which describes the dependence structure between the variables (how width correlates with height). The marginal distributions of width and heights of humans are roughly Gaussian, but sampling from these marginals is not satisfactory, since

By using a positive correlated copula : by using a positively correlated copula, one can make sure to sample the variables in a way where the probability of river flooding does increase with maximum river level.

In a Gaussian copula, the dependency between random variables is assumed Gaussian. For Gaussian copulas, the dependence structure is that of a Gaussian multivariate which is entirely dictated by the associated covariance matrix.

## 2.6 Hierarchical modeling

# 3 Modeling the effect of strategy on endpoints

We first investigate the effect of participant strategy on movement time and  $ID_e$ , using the Guiard-Olafsdottir (GO) dataset [?].

## 3.1 Summary of the paradigm used to produce the GO dataset

The pointing task used to produce the GO dataset is not a typical Fitts task since it has no predefined width. It is thus a pointing task without nominal accuracy, where participants aim towards a point in space while self-regulating their accuracy, in line with experimenter instructions. These instructions go from speed emphasis to precision emphasis in 5 steps (speed emphasis, speed, balanced, precision, precision emphasis). In speed emphasis, the participant is instructed to go as fast as possible while disregarding accuracy but to make sure that on average the endpoints are centered on the target *i.e.*, large under and overshoots are admitted, as long as they are about as equally probable. In precision emphasis, the participant is instructed to hit the (1 pixel) target, without caring for movement time. In the balanced

condition, participants are instructed to adopt a *natural* strategy *i.e.*, the one they would have adopted without explicit instructions. For the speed and precision instructions, the participants should behave in between the balanced and emphasis conditions.

While in such a paradigm an ID can not be defined a priori, the standard deviation of the resulting endpoints can be computed, which means an  $ID_e$  can be computed a posteriori. Previous work has shown that data produced this way is in line with data from a classical Fitts' law [?, ?, ?, ?], in particular, the mean movement time  $\overline{MT}$  correlates strongly with  $ID_e$ .

### 3.2 The GO dataset

A summary of the GO dataset obtained by plotting  $\overline{MT}$  against  $ID_e$  for all participants, clustered by strategy, is shown Figure 1. Participants are able to produce movements with a wide range of  $ID_e$ , from 1 to 9 bit, depending on the strategy. The different strategies are quite distinct from each other, which implies that participants were able to correctly self-regulate their strategies, and that this self-regulation was quite consistent across different participants.

We find that  $\overline{MT}$  increases with  $ID_e$  as expected. This is true both between strategies, where it is clear that emphasizing accuracy leads to larger value of  $ID_e$  and  $\overline{MT}$ , but also within a particular strategy as indicated by the orientation of the 95% prediction ellipses<sup>1</sup> that result from a bivariate Gaussian fit to each strategy.

### 3.3 A bivariate Gaussian model for $\overline{MT}$ and $ID_e$ per strategy

Since the data points form correlated cluster, we model the joint increase of  $\overline{MT}$  and  $ID_e$  with a bivariate Gaussian model, for each of the five strategies of the GO dataset. We chose the bivariate model because it is simple, easy to interpret, expresses the dependence between  $\overline{MT}$  and  $ID_e$  via correlation, and is expected to fit reasonably well. The fit for each strategy is displayed by drawing the associated 95% prediction ellipsis in Figure 1.

$$\begin{pmatrix} ID_e \\ \overline{MT} \end{pmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_i \\ \mu_t \end{bmatrix}, \begin{bmatrix} \sigma_i^2 & \rho\sigma_i\sigma_t \\ \rho\sigma_i\sigma_t & \sigma_t^2 \end{bmatrix} \right). \quad (13)$$

---

<sup>1</sup>The covariance matrix of a Gaussian bivariate is the rightmost term of Equation 13 where  $\rho$  is the correlation between the two components. The angle of the prediction ellipsis is given by  $\frac{1}{2} \arctan \left( \frac{2\rho\sigma_i\sigma_t}{\sigma_i^2 - \sigma_t^2} \right)$ .

It is quite clear that when the strategy emphasizes precision, the mean of the Gaussian vector increases, as well as the size of the prediction ellipsis; however, its orientation stays relatively constant, suggesting that the correlation between the  $ID_e$  and  $\overline{MT}$  may also be.



Figure 1: The effective Fitts' law fitted on data from the GO dataset. Data is pooled for all participants, and separated by strategies: from emphasis on speed to emphasis on accuracy. Each point corresponds to a block aggregate. A bivariate Gaussian fit, computed for all points of the same strategy, is also displayed through the associated 95% prediction ellipsis.

### 3.4 A bivariate Gaussian model for $\overline{MT}$ and $ID_e$ for different strategies

In the GO dataset, strategy is an ordinal variable: we can order the strategies but can not quantify the *distance* between two strategies. In the interest of a more useful quantitative model, we'd like to consider strategy as a numerical

variable instead. Considering the five instructions cover the full range of available strategies, a naïve approach is to map the five strategies to equally spaced points on  $[-1,1]$ , where the balanced strategy maps to 0, and 1 (resp. -1) maps to speed emphasis (resp. precision emphasis).

We then fitted a linear model on the mean and covariance components (Equation 13) previously identified against the “numerical strategies”. The fits are displayed Figure 2; the full fit summaries can be found in the supplementary materials. We observe that

- The linear increase of the mean with numerical strategy is clear,
- The increase of  $\sigma_i$  and  $\sigma_t$  with numerical strategy can also reasonably be described by a linear model, although evidence is less strong than for the mean, especially for  $\sigma_i$ ,
- the correlation between  $\overline{MT}$  and  $ID_e$  remains relatively constant across strategies, with a mean value of  $\rho = 0.36$ .

These results suggest a simple model, where the mean and standard deviations of  $ID_e$  and  $\overline{MT}$  increase linearly with strategy, and where  $ID_e$  and  $\overline{MT}$  remain moderately correlated ( $\rho = 0.36$ ) throughout strategies in the GO dataset.

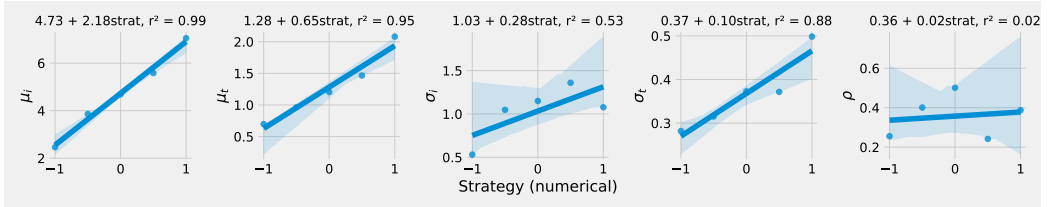


Figure 2: The parameters of the mean and covariance of the bivariate Gaussian model defined in Equation 13 plotted against the five numerical strategies. The associated linear fit is displayed above each parameter, together with a measure of goodness of fit.

## 4 Modeling variability of the EMG parameters at the population level

In this section, we use an existing dataset by Jude Guinness and Poor [?] (JGP) to model variability of the parameter values of the EMG model fitted



to pointing data at the population level. The JGP dataset was produced with the well-known 2D multi-directional tapping task [?], using the FittsStudy software [?]. The experiment was replicated 6 times (on three days, twice a day).

We fitted the EMG model to data from the JGP dataset pooled by participant using maximum likelihood estimation, with `emgregs` [?], a Python library for inference (and simulation) of EMG models. We first confirmed the expected result that the EMG model fits much better than the Gaussian model, with the lowest difference in AIC of more than 250<sup>2</sup>. EMG fits and AIC differences with the Gaussian model for each participant can be found in the supplementary materials.

## 4.1 Population variability

We then described how the estimated parameters of the EMG distribution vary across the population. The EMG distribution used Equation 9 has five parameters:  $\beta = [\beta_0, \beta_1]$ , the two parameters of the linear fit Equation 5,  $\sigma$ , the symmetric noise component and  $\lambda = [\lambda_0, \lambda_1]$ , the two asymmetric noise components. As in the previous section, we could jointly describe the probability distribution of these five parameters, the problem is that this requires describing the marginal distributions as well as their dependences, which signifies an increase in parameters. For example, a multivariate Gaussian would need  $5 + 15 = 20$  parameters to be estimated, for which there are too few participants in the JGP dataset. We therefore made the following assumptions to limit the overall number of parameters to estimate:

- we assume  $\sigma$  is independent of the other parameters, because the interpretation of  $\sigma$  in the EMG distribution is as producing noise accounting for modeling errors and random departures.
- We assume that  $\beta_0$  and  $\beta_1$  are dependent. Two reasons why  $\beta_0$  may correlate with  $\beta_1$  are: 1) some data points may have the effect of tilting the linear fit (*e.g.*, a few data points of high  $ID_e$  and low MT may drag the slope down, but considering the lower ID data remains unchanged, the fit will pivot around low ID values, which mechanically increases the intercept) 2) an overall “slow” participant may likely have larger intercept and larger slope. However, we assume  $\beta$  to be independent of other parameters: the best performing data is unrelated to the variability of the data, mostly produced by  $\lambda$ .

---

<sup>2</sup>Typically, a difference in AIC of 10 is considered good support to reject a candidate model, so consistent AIC differences of more than 250 gives plenty of reason here to reject the Gaussian model.

- We assume that  $\lambda_0$  and  $\lambda_1$  are correlated, by lack of better knowledge.

The estimated EMG parameters are represented Figure 3. The left panel shows the two  $\beta$  parameters. We notice a negative correlation (increasing the slope is associated with decreasing the intercept), in line with the “tilting” explanation. A bivariate Gaussian fit, represented by a 95% prediction interval, seems reasonable. The middle left panel shows a histogram of  $\sigma$ . A reasonable model here seems to be a uniform distribution. The middle right panel shows that  $\lambda_0$  and  $\lambda_1$  do not seem to correlate nor can they be modeled reasonably by a Gaussian bivariate. Furthermore,  $\lambda_0$  frequently is identified with value 0. For completeness, we also investigate the distribution of  $\lambda_1$ , which can be well fit by a Gaussian. As a result, we adopt the following population model for the EMG parameters:

$$\beta \sim \mathcal{N}\left(\begin{bmatrix} 0.08 \\ 0.21 \end{bmatrix}, \begin{bmatrix} 0.003 & -0.001 \\ -0.001 & 0.001 \end{bmatrix}\right) \quad (14)$$

$$\sigma \sim \mathcal{U}[0.04, 0.09] \quad (15)$$

$$\lambda_0 = 0.05 \quad (16)$$

$$\lambda_1 \sim \mathcal{N}(0.1, 0.02) \quad (17)$$

$$(18)$$

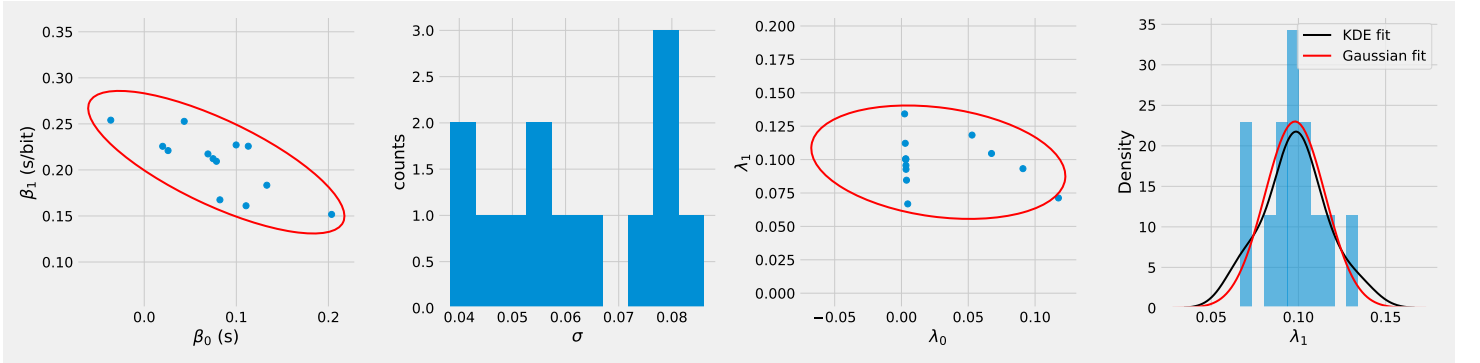


Figure 3: Parameters of the EMG fits as defined Equation 9 on the JGP dataset. Left:  $\beta$  fit with a bivariate Gaussian. Middle left: histogram of  $\sigma$ . Middle right: bivariate Gaussian fit to  $\lambda_0$  and  $\lambda_1$ . Right: Density plot (normalized histogram), with kernel density estimate (KDE, black) and Gaussian (red) fits.

## 5 Combining EMG and effect of strategy: candidate models

In section 3 we showed how to relate a strategy parameter to  $p(\text{ID}_e, \overline{\text{MT}})$ , the joint distribution of  $\text{ID}_e$  and  $\overline{\text{MT}}$ . In section 4, we described  $p(\text{MT}|\text{ID}_e)$ , the conditional distribution of MT knowing  $\text{ID}_e$ , via the EMG model at the population level. Our challenge is to describe a procedure to *jointly* sample  $\text{ID}_e$  and MT such that both aforementioned results remain valid. We propose several procedures.

**Sampling from  $\text{ID}_e$**  The first idea is to sample  $\text{ID}_e$  from a Gaussian distribution, and to plug it directly in the EMG. The reason this will work is that the EMG distribution naturally correlates  $\text{ID}_e$  with the mean MT see Appendix A. The sampling scheme is then quite simple:

1. Sample  $x$  from the marginal  $\text{ID}_e$  distribution

$$x \sim \mathcal{N}(a + bs, (c + ds)^2), \quad (19)$$

where  $a$ ,  $b$ ,  $c$  and  $d$  are chosen by the modeler *e.g.*, estimated from a dataset as in Figure 2.

2. Sample  $y|x$  from the  $EMG(\beta x, \sigma^2 \lambda x)$  distribution where  $\beta$ ,  $\sigma$ ,  $\lambda$  are chosen by the modeler *e.g.*, estimated from a dataset as in Figure 3.

**Jointly sampling  $\text{ID}_e$  and the mean MT** The second idea is to directly sample from the jointly gaussian model Equation 13. This guarantees that the values of  $\text{ID}_e$  and the mean of MT are distributed as indicated ???. The  $\beta$  parameter remains unchanged since it should remain invariant [?, ?], but  $\lambda$  is adapted so that the inferred mean of the EMG does not change with regards to the sampled  $\mu(\text{MT})$ . The sampling scheme is

1. Sample  $(\mu(x), y)$  from the joint distribution Equation 13
2. Sample  $y|x$  from the  $EMG(\beta x, \sigma, \lambda x)$  distribution where

$$\lambda_1 = (\mu(\text{MT}) - \lambda_0 - \beta_0 - \beta_1 x)/x \quad (20)$$

**Gaussian copulas** to do:

- make a proper library with the code for gen data -i fix copulas with copulas.py

- generate three figures with datasets and diagnostics
- multivariate kolmogorov smirnov test to compare empirical cdfs between two datasets <https://cran.r-project.org/web/packages/fasano.franceschini.test/fasano.franceschini.test.pdf>
- test on throughput – take a known dataset, compute throughput and estimate standard dev with bootstrap. Do the same with throughput on generated data.

**Hierarchical model** For a hierarchical model, one can further sample  $\beta$ ,  $\sigma$  and  $\lambda$  from distributions, according to Figure 3:

$$\beta \sim \mathcal{N}(\mu, \Sigma) \quad (21)$$

$$\sigma \sim \mathcal{U}(u_{\min}, u_{\max}) \quad (22)$$

$$\lambda_0 = 0.05 \quad (23)$$

$$\lambda_1 \sim \mathcal{N}(0.1, 0.02) \quad (24)$$

## 6 Extra

### 6.1 Deciding on the strategy with $\beta$ and the WHO model

### 6.2 Modeling participant strategy with the Linear Quadratic Regulator (LQR)

\*\*likely remove because there is no stopping point operationalized in the LQR/LQG.\*\*

One of the simplest optimal control setting is the so-called Linear Quadratic Regulator (LQR). The LQR assumes that the dynamics are linear

$$\dot{x}(t) = f(x(t), u(t)) = Ax(t) + Bu(t), \quad (25)$$

where  $A$  and  $B$  are two matrices of proper dimension, and the cost function is a quadratic function<sup>3</sup>,

$$\mathcal{J}_T = \int_T^f x(t)^t Q x(t) + u(t)^t R u(t), \quad (26)$$

---

<sup>3</sup>We use here a loose notation for the summation for simplicity, since we have not specified whether time is continuous or discrete. Solutions of the problem will usually depend on the horizon (duration of  $T$ , which can be  $\infty$ ), and whether signals are continuous or discrete time. We refer the interested reader to the supplementary materials for more information.

where  $Q$  and  $R$  are positive semi-definite matrices of proper dimension. With these assumptions, it becomes possible to find a closed-form solution is that one can compute the  $u(t)$  that minimizes an arbitrary but *quadratic* cost function of the state and command

$$\mathcal{J} = \int_T x^t Q x + u^t R u, \quad (27)$$

where  $R$  and  $Q$  are two matrices that indicate costs relative to the value of the state  $x(t)$  and the value of the control signal  $u(t)$ . This model can be made more complex to better fit human behavior:

## A Proof of Result 1

In general, and denoting the speed-accuracy strategy as  $s$ , the sampling scheme is

$$\text{ID}_e \sim \mathcal{N}(a + b s, (c + d s)^2), \quad (28)$$

where  $a$ ,  $b$ ,  $c$  and  $d$  are parameters determined in Figure 2. In general, the mean  $\mu(\text{EMG}(\beta, \sigma, \lambda))$  and variance  $\text{Var}(\text{EMG}(\beta, \sigma, \lambda))$  of an EMG distribution with parameters  $(\beta, \sigma, \lambda)$  is given by

$$\mu(\text{EMG}(\beta, \sigma, \lambda)) = \beta + \lambda \quad (29)$$

$$\text{Var}(\text{EMG}(\beta, \sigma, \lambda)) = \sigma^2 + \lambda^2 \quad (30)$$

With  $y$  the movement time and  $x$  the effective index of difficulty  $\text{ID}_e$ , we have

$$\mathbb{E}_x[\mu(y|x)] = \mathbb{E}_x[\beta_0 + \beta_1 x + \lambda_0 + \lambda_1 x] = \beta_0 + \lambda_0 + (\beta_1 + \lambda_1)\mathbb{E}[x] \quad (31)$$

$$= \beta_0 + \lambda_0 + (\beta_1 + \lambda_1)a + (\beta_1 + \lambda_1)bs \quad (32)$$

$$= a' + b's, \quad (33)$$

which shows the mean value of MT to scale linearly with  $s$ . For the variance of MT, we can perform a similar computation and see it remains a second order polynomial in  $s$

$$\mathbb{E}_x[\text{Var}[y|x]] = \mathbb{E}_x[\sigma^2 + (\lambda x)^2] = \sigma^2 + \mathbb{E}_x[(\lambda_0 + \lambda_1 x)^2] \quad (34)$$

$$= \sigma^2 + \lambda_0^2 + 2\lambda_0\lambda_1\mathbb{E}_x[x] + \lambda_1^2\mathbb{E}[x^2] \quad (35)$$

$$= \sigma^2 + \lambda_0^2 + 2\lambda_0\lambda_1(a + bs) + \lambda_1^2 [\text{Var}(x) + \mu(x)^2] \quad (36)$$

$$= \sigma^2 + \lambda_0^2 + 2\lambda_0\lambda_1(a + bs) + \lambda_1^2 [(c + ds)^2 + (a + bs)^2] \quad (37)$$

$$= p_0 + p_1 s + p_2 s^2 \quad (38)$$