One Fitts' Law, Two Metrics

Julien Gori[1(✉)], Olivier Rioul[1], Yves Guiard[1,2], and Michel Beaudouin-Lafon[2]

[1] LTCI, Telecom ParisTech, Université Paris-Saclay, 75013 Paris, France
{julien.gori,olivier.rioul,yves.guiard}@telecom-paristech.fr
[2] LRI, Univ. Paris-Sud, CNRS, Inria, Université Paris-Saclay, 91400 Orsay, France
mbl@lri.fr

Abstract. Movement time in Fitts' law is usually considered through the ambiguous notion of the average of minimum movement times. In this paper, we argue that using two distinct metrics, one relating to minimum time and the other relating to average time can be advantageous. Both metrics have a lot of support from theoretical and empirical perspectives. We also give two examples, one in a controlled experiment and the other in a field study of pointing, where making the minimum versus average distinction is fruitful.

1 Introduction

In Human Computer Interaction (HCI), Fitts' law is recognized as "the law of pointing", and is regularly used for, e.g., device evaluation or interface design. It relates the time MT to reach a target to target size W and target distance D. The law is described by the following equation:

$$\mathrm{MT} = a + b \cdot \mathrm{ID}, \quad (1)$$

where MT represents movement time and ID represents a function of D and W. Most of the efforts of researchers, since Fitts' seminal paper [5], have been directed towards the study of ID; initially Fitts [5] proposed $\mathrm{ID} = \log_2 \frac{2D}{W}$, but today the most common formulation in HCI is the so-called Shannon formulation, due to MacKenzie [11], $\mathrm{ID} = \log_2 \left(1 + \frac{D}{W}\right)$. Perhaps surprisingly, the interpretation of MT has rarely been discussed, yet there is room for doubt: is the measure under consideration a minimum, an average, or an average minimum?

1.1 Time Metrics

Fitts' formal hypothesis in 1954 [5] was that: *"If the amplitude and tolerance limits of a task are controlled by E [the experimenter], and S [the participant] is instructed to work at his maximum rate, then the average time per response will be directly proportional to the minimum amount of information per response demanded by the particular conditions of amplitude and tolerance"* [5, p. 2].

R. Bernhaupt et al. (Eds.): INTERACT 2017, Part III, LNCS 10515, pp. 525–533, 2017.
DOI: 10.1007/978-3-319-67687-6_36

Note that the expression *minimum amount of information* is used by Fitts to denote ID, which is not the matter here (see, e.g., [13] for an analysis of ID). For movement time Fitts gave the formula $\text{MT} = \frac{\text{ID}}{C}$, where C is the capacity of the human motor channel, representing the participant's *maximum rate*. If C is a maximum, the movement times should correspond to minimum movement times. Yet, we should, according to Fitts, consider the average of those minima over all time measures. Fitts' MT metric is thus an average of the minimum movement times. This raises three issues:

– Evidently, the oxymoron "average-minimum" described by Fitts is confusing. Several authors use, instead, different wordings, which suggest other interpretations of MT. A few examples are given by Soukoreff et al. [13] who consider "movement time performance for rapid aimed movements", Hoffman [8] "movement time", and Drewes [3] "mean time".
– Fitts needs the participant (S) to work at his or her maximum rate, so that the resulting movement times reflect S's full commitment to the pointing task. Yet, it is well documented that subjects are rarely fully committed to boring, repetitive tasks such as those composing Fitts' pointing experiments. The use of the average of MT's is thus based on a false premise, resulting from a shortcoming of the experimental design. It is unreasonable to take it for granted that the participants routinely perform to their fullest capabilities during the course of an experiment, even if they were instructed to do so.
– Fitts' average-minimum metric cannot be defined in a field study, because the participant are never instructed to operate at their maximum rate.

In this paper, we argue that there is a lot to be gained if one separates the average-minimum metric into two distinct metrics instead: a *minimum time metric* that relates to *extreme* performance, and an *average time metric* that relates to *mean* performance; each metric bringing its own valuable information. We show that there is enough evidence in Fitts' law literature that both follow Fitts' law, albeit with different slopes and intercepts, therefore providing essentially two Fitts' laws for the price of one.

2 Fitts' Law Is About Extreme Performance

We have two reasons to assert that Fitts' law should be interpreted as a law of extreme performance: First, a theoretical argument that exploits and precisely interprets the information theoretic analogy that was first used by Fitts, and subsequently successfully considered by, e.g., MacKenzie [9,11] and Zhai et al. [15]; Second, a pragmatic argument related to the recent unification of Schmidt's and Fitts' laws, due to Guiard and Rioul [7].

2.1 Shannon's Channel Capacity Theorem

Fitts derived his law by analogy with Shannon's channel capacity for the Gaussian additive channel. In information theory, the capacity of a channel of input X and output Y is defined by $C = \max_{p(x)} I(X;Y)$, where $I(X;Y)$ is the mutual information between X and Y and the maximum is taken over all input probability distributions. The channel coding theorem [2] states that C is the *maximum allowable bit rate* for which reliable transmission is possible over the considered channel. The capacity is reached at the limit of a (perfectly) optimal coding scheme. Anything less will give lower transmitted information.

2.2 Unified Pointing Paradigms

Recently, Guiard and Rioul [7] have shown that the time minimisation paradigm introduced by Fitts [5], the spread minimisation paradigm of Schmidt et al. [12] and the dual minimisation paradigm of Guiard [6] can receive a unified account provided that the participants are assumed to invest less than 100 percent of their resource in their performance. Accordingly, only the best performing samples should be expected to describe the speed-accuracy trade off. The less-than-total resource investment assumption thus matches common sense, as well as the information theoretic concept of capacity. Based on this idea, Guiard and Rioul were able to merge the linear law of Schmidt and the logarithmic description of Fitts law, describing them as different regions of the same speed-accuracy trade-off function.

Thus, Fitts' law should estimate an extremum, rather than an average. In other words:

- for a set of movements that have the same duration, only the movements with the highest ID are susceptible of achieving the capacity, or equivalently,
- for a set of movements of fixed ID, only the movements of shortest duration are susceptible of achieving the capacity.

This is in line with the idea that Fitts' law can only be seen as an extreme performance in a constrained condition of speed. This rationale makes sense: We cannot expect a model to account for the perturbations induced by a badly designed experiment, fatigue, or lassitude.

3 Two Metrics

3.1 The Average Time Metric

The average time metric is defined through the following relation:

$$\mu(\mathrm{MT}) = a + b \cdot \mu(\mathrm{ID}), \tag{2}$$

where $\mu(X)$ is the mean of X per block. In controlled experiments with speeding instructions, $\mu(\mathrm{ID}) = \mathrm{ID}$. Controlled Fitts' law studies exclusively handle the average metric (see, e.g., [13], which is used as a reference by many researchers). There is thus an abundance of results showing that this linear relation between mean MT and mean ID holds well and provides for good fits. Field studies such as Chapuis et al. [1] have shown that this relationship would hold as well, with r^2

coefficients for mean movement times regularly close to 1 (min 0.740, max 0.988, mean 0.941, std. dev. 0.044 for 24 different participants). The usual method to fit the average law is to perform linear regression on the data averaged per block (e.g. [13]).

3.2 The Minimum Time Metric

The minimum time metric for Fitts' law is defined as:

$$\min(\mathrm{MT}) = a + b \cdot \mathrm{ID}, \tag{3}$$

in line with the theoretical arguments of the previous section. We will now call this relationship Fitts' *min* metric. We next give an illustration of the min metric from data acquired in the field.

3.3 A Field Study Example

Figure 1 illustrate Fitts' min and average metrics, using a field study by Chapuis et al. [1]. For several months, Chapuis et al. unobtrusively logged cursor motion from several participants using their own hardware. They were able to identify offline the start and end of movements as well as the target information for several hundreds of thousands of click-terminated movements. With this information, one can then represent the movements in a MT versus ID graph, as normally done in a controlled Fitts law experiment.

Chapuis and his colleagues kindly allowed us to access their raw data. To compute task difficulty in the 2D space of computer screens we followed the suggestion given by Mackenzie and Buxton [10] and chose $\mathrm{ID} = \log_2\left(1 + \frac{D}{\min(H,W)}\right)$, where H and W are the height and width of the target, respectively. Whenever an item was clicked, it was considered the target, hence target misses were not considered as errors in [1].

Figure 1 shows the data from one representative participant (P3) of Chapuis et al. 's field study. The I_D axis is truncated at 6 bits because beyond that level of difficulty the density of data points dropped dramatically.

Obviously, the data obtained with no speeding instructions (and no experimenter to recall them) exhibits a huge amount of stochastic variability along both dimensions of the plot. While in the X dimension, most I_D values fell in the range from 0.5 to 6 bits (presumably a reflection of the geometric layout of the graphical user interface), the variability along the Y dimension is extreme.

Linear regression performed on the whole data set of Fig. 1 shows that in the field experiment of Chapuis et al. movement time varied totally independently of the index of difficulty. With an r-square coefficient close to 0 ($r^2 = 0.03$), this data seems to totally fail to corroborate Fitts' law. This impression, however, is quite false.

In the right panel of Fig. 1, which ignores all MT data above 2 s and thus zooms-in on the Y-axis towards the bottom of the plot, one can distinctly see that the bottom edge of the cloud of data points does not touch the X axis.

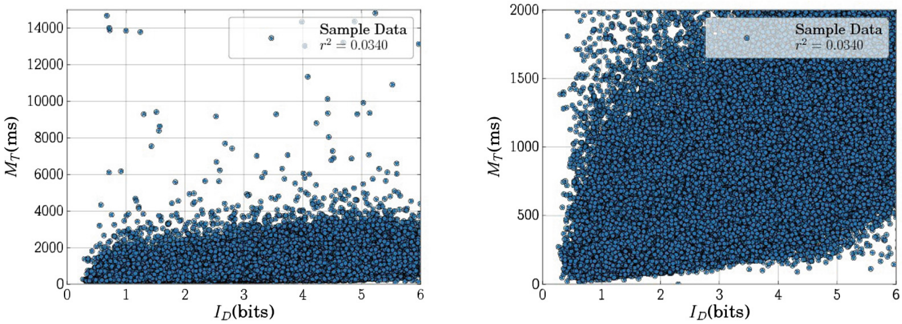


Fig. 1. Movement time as a function of task difficulty in one representative participant of [1]. Shown are over 90,000 individual movement measures. *Left*: MT up to 15 s. *Right*: MT up to 2 s. (lower part of the graph on the left)

Rather, in the downward direction, the density drops sharply: no matter the ID region considered, the distribution of performance measures has an unending tail above and what we call a *front* below, which in comparison is very steep (see [1] for the histogram of movement times). The unending tail is understandable as "it is always possible to do worse" [7]. In contrast, the movement time cannot be reduced below a certain critical value which accurately defines the front.

A closer look at the right panel of Fig. 1 reveals that the bottom edge of the scatter plot is approximately *linear*: this linear edge is what justifies Fitts' law. In other words, the empirical regularity in Fitts' law is, in essence, a *front of performance*, a lower bound that cannot be passed by human performance. Figure 1 also reveals a number of presumable fast outliers. Many reasons may explain why a small proportion of data cross the frontier. Some data points may correspond to lucky movements, failures in the analysis software, etc.

A front of performance is observable in data from the field study of Chapuis et al. because unsupervised everyday pointing does offer, albeit in a minority of cases, opportunities to perform with high levels of speed and accuracy. The difference between a field study and a controlled experiment is thus one of degree, not of nature. Experimenters have recourse to pressurizing speed/accuracy instructions simply to get rid of endless, uninformative, tails in their distributions of MT measures.

3.4 Estimating the Parameters of Fitts' Law Min

Estimates of parameters a and b of Eq. 3 cannot be obtained through conventional linear regression, which leverages all movement-time measures. We tentatively designed a new technique that estimates a and b for Fitts' min metric. First, slice the MT vs. ID scatter plot into vertical slices (i.e. split the range of ID into contiguous intervals). Then, for each slice (i.e. for all MT measures that fall inside each ID interval):

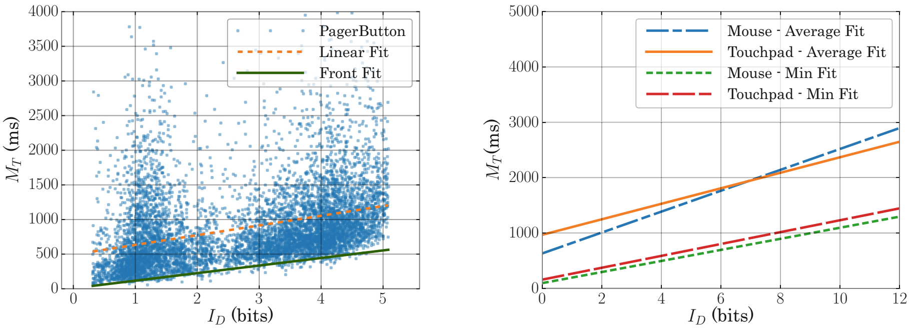
1. Draw the frequency histogram of MT for the interval (the histogram will exhibit strong positive skew, with its mode close to its minimum value and with an extended tail);
2. Fit the portion between the mode and the minimum value of each distribution with a linear curve (we observed that distributions fronts are not just very steep, more often than not they are almost linear);
3. Find the intercept of this curve: this is an estimate of the minimum value of MT for the ID interval considered.

Finally, plot these intercepts as a function of the ID and find the line that best fits the new scatter plot. The final step is to estimate the intercept a and the slope b of Eq. 3.

The technique has the following characteristics:

1. It specifies the parameters of a straight line, and this is Fitts' min in its law form;
2. It takes into account only the points corresponding to the best performance, through the exclusive consideration of the fronts of the performance distributions. The technique does not just discard "slow" outliers—i.e. movements of unusually long duration—it ignores all data points, but the very best.
3. It also eliminates "fast" outliers, i.e. movements of abnormally short duration.

Figure 2(a) displays the min law obtained for one participant pointing at a pager button[1]. Intercepts and slopes have been computed using both linear regression and best-performance fit; the parameters are respectively $a = 804$ ms and $\frac{1}{b} = 9.34$ bit/s and $a = 79$ ms and $\frac{1}{b} = 8.20$ bit/s. The slopes are similar



(a) (b)

Fig. 2. (a) Fitting Fitts' law to the pager button data-set for a single participant. Red line: linear regression; blue line: front of performance. (b) Comparison between Fitts' min law versus Fitts' average law, for one participant using mouse and touch-pad. (Color figure online)

[1] A button such as "Next Page" and "Previous Page" in a paginated control.

but the intercepts are strikingly different. While the best-performance fit, corresponding to the min law, yields intercepts close to those met in controlled experiments, linear regression, corresponding to the average law, produces very high intercepts, comparable to those reported by Evans and Wobbrock [4].

4 Using the Two Metrics

Having two metrics instead of one can only add precision in the analysis of movements, and allows to observe phenomena that were not possible before. We give two examples, one for controlled experiments and a second for field studies, to illustrate the potential benefits.

4.1 A Possible Scenario for Controlled Experiments

As already emphasized, a Fitts task requires participants to work at their maximum rate. The quality of the measures depends on the experimenter's success at motivating participants. With participants that performed extremely well, indeed the min law and the average law would be extremely close to each other. An easy way to measure that is simply to quantify the difference between the two fits. The closer they are, the better the experiment. The distance between the two fits is thus in this particular case a measure of the quality of the experiment, and could prove well suited as a quantifier of experimental noise.

4.2 An Example from a Field Study

Field studies of pointing are very useful in the sense that they give us information about the natural behavior of users in the real world. The variability of movement times will be much higher than in a controlled experiments. The min and the average variants of Fitts' laws thus provide two very different types of information, the former giving the parameters of the law for best performance, and the latter giving average performance. The min law is expected to remain unchanged across systems (e.g., which operating system), environmental conditions (e.g., in a noisy crowded room or on the contrary alone in a well isolated office), human conditions (e.g., tired and bored or time-pressured), whereas the average law is expected to reflect all these changes. Importantly, the min law is expected to give similar results for both controlled experiments or field studies, whereas the average law is a reflection of all the differences that can occur between a controlled experiment and a field study.

As an example, we provide a comparison between mouse and touch-pad from Chapuis et al.'s dataset. We compared the behavior of one participant who happened to regularly alternate between the mouse and the touch-pad. The results of the fits for each device are displayed in Fig. 2(b), and the parameters are reported in Table 1.

With linear regression, the touch-pad has a higher intercept than the mouse, but the mouse has a steeper slope, hence the lines cross at about ID = 7.

Table 1. Intercept (a) and inverse of slope ($\frac{1}{b}$) of the min and average law for the mouse and touch-pad for participant 3 from Chapuis et al.'s experiment [1].

	Intercept (ms)		$\frac{1}{b}$ (bit per s)	
	Average	Min	Average	Min
Mouse	647	114	5.46	10.81
Touch-pad	968	178	7.15	10.23

This would suggest that the touch-pad is slower than the mouse for most tasks, but is faster for high-difficulty tasks. This observation is untypical and needs investigation: is it that the effort required to handle the touch-pad is less than for the mouse, or is the average law victim of some artefact, such as irregular distribution of ID or unusually long movement times for low ID, which tilts the linear regression fit? By contrast, our performance fit tends to show that the mouse is always a bit faster than the touch-pad. This result is consistent with those found in the literature for controlled experiments, e.g., Yun and Lee [14].

5 Conclusion

There are good reasons to use two metrics for Fitts' law. On the one hand, the min metric has a solid theoretical background, supported by Shannon's channel coding theorem and Guiard & Rioul's study of the speed-accuracy tradeoff. On the other hand, the average law can claim more than 60 years of empirical validation. We expect the min metric to give robust and consistent results. It relates to the human motor capacity, and hence is a useful reference for HCI experimenters. The average metric is expected to reflect everything that is not related to the human motor system, e.g. environment or device.

We have given two examples where the min metric provides useful information when compared to the average metric. In the case of the field study of mouse versus touchpad, the min law gives results that are consistent for both field study and controlled experiment. We expect this to be a general property. For the average law, the touchpad apparently outperforms the mouse for high ID's. This is probably worth investigating. This dual characterization certainly opens new perspectives for Fitts' law in HCI research, in both field studies and controlled experiments, and may lead to finer results.

References

1. Chapuis, O., Blanch, R., Beaudouin-Lafon, M.: Fitts' law in the wild: a field study of aimed movements (2007)
2. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New Jersey (2012)
3. Drewes, H.: Only one Fitts' law formula please! In: Extended Abstracts on Human Factors in Computing Systems, CHI 2010, pp. 2813–2822. ACM (2010)

4. Evans, A., Wobbrock, J.: Taming wild behavior: the input observer for obtaining text entry and mouse pointing measures from everyday computer use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2012, NY, USA, pp. 1947–1956 (2012). http://doi.acm.org/10.1145/2207676.2208338
5. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol. **47**(6), 381 (1954)
6. Guiard, Y., Olafsdottir, H.B., Perrault, S.T.: Fitt's law as an explicit time/error trade-off. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1619–1628. ACM (2011)
7. Guiard, Y., Rioul, O.: A mathematical description of the speed/accuracy trade-off of aimed movement. In: Proceedings of the 2015 British HCI Conference, pp. 91–100. ACM (2015)
8. Hoffmann, E.R.: Which version/variation of Fitts law? A critique of information-theory models. J. Mot. Behav. **45**(3), 205–215 (2013)
9. Mackenzie, I.S.: Fitts' law as a performance model in human-computer interaction. Ph.D. thesis, University of Toronto (1992)
10. MacKenzie, I.S., Buxton, W.: Extending Fitts' law to two-dimensional tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 219–226. ACM, New York (1992)
11. MacKenzie, I.S.: A note on the information-theoretic basis for Fitts law. J. Mot. Behav. **21**(3), 323–330 (1989)
12. Schmidt, R.A., Zelaznik, H., Hawkins, B., Frank, J.S., Quinn Jr., J.T.: Motor-output variability: a theory for the accuracy of rapid motor acts. Psychol. Rev. **86**(5), 415 (1979)
13. Soukoreff, R.W., MacKenzie, I.S.: 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 (2004)
14. Yun, S., Lee, G.: Design and comparison of acceleration methods for touchpad. In: Extended Abstracts on Human Factors in Computing Systems, CHI 2007, CHI EA 2007, pp. 2801–2812. ACM, New York (2007). http://doi.acm.org/10.1145/1240866.1241082
15. Zhai, S., Kristensson, P.O., Appert, C., Andersen, T.H., Cao, X.: Foundational issues in touch-screen stroke gesture design-an integrative review. Found. Trends Hum.-Comput. Interact. **5**(2), 97–205 (2012)