To Miss is Human: Information-Theoretic Rationale for Target Misses in Fitts' Law

Julien Gori, Olivier Rioul and Yves Guiard
LTCI, CNRS, Telecom ParisTech
Université Paris-Saclay
Paris, France
firstname.lastname@telecom-paristech.fr

ABSTRACT
In usual Fitts' law experiments the outcome of a pointing act can be either measured as an error, i.e., a distance from endpoint to target center, or categorized in an all-or-none way as a hit versus a miss. Information theory offers a useful distinction between transmission errors (the received symbol is wrong) and erasures (the received symbol is empty). Although Fitts' law research has been very much inspired by the information theoretic rationale, the error/erasure distinction has escaped attention so far: Target misses have always been treated as normally-distributed errors, through the effective index of difficulty ID_e. The paper introduces a new index of difficulty based on the simple observation that a target miss conveys zero bit of information, i.e., it is an erasure. Not only is the new index more consistent with the fundamentals of information theory, it is much simpler to derive than the ISO-recommended ID_e.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces; Theory and Methods

Author Keywords
Information Theory; Fitts' Law; Human Performance Models

INTRODUCTION
Fitts' law is a well documented rule [13] that predicts the time MT it takes to reach a target of width W located at distance D. The law conveniently reduces the two-parameter task (D,W) to a single parameter called the index of difficulty ID [12]:

$$ID = \log_2\left(1+\frac{D}{W}\right) \qquad \text{bit}^1.$$

[1] This form of the ID is known as the Shannon ID [11]. Other well-known forms are the Welford $ID = \log_2(\frac{1}{2}+\frac{D}{W})$ [18] and the Fitts $ID = \log_2(\frac{2D}{W})$ [4].

CHI 2017, May 6-11, 2017, Denver, CO, USA.
Copyright © 2017 ACM ISBN 978-1-4503-4655-9/17/05 ...$15.00.
http://dx.doi.org/10.1145/3025453.3025660

Movement time MT has been found empirically to be linearly related to ID:

$$MT = a + b \cdot ID. \qquad (1)$$

The experimental paradigm introduced by Fitts [4] consists of measuring MT, which participants try their best to minimize, at various levels of ID. Ideally, movement endpoints must always fall within the prescribed tolerance interval of width W—or at least with very few misses. Unfortunately, the paradigm faces a serious difficulty: As has been observed ever since Fitts [4], the percentage of target misses tends to increase systematically, along with MT, as ID is raised, meaning that experimenters face two effects of the same manipulation, which is one too many. Different participants may also have different biases towards speed or accuracy (see e.g. [6] for further discussion on the innate bias in psycho-physical experiments).

The currently received—and actually standardized [10, 17]—solution is to adjust the ID by computing an *effective* index of difficulty ID_e in the place of the prescribed or *nominal* ID. The adjusted index takes into account the empirical standard deviation σ of the endpoint distribution:

$$ID_e = \log_2\left(1+\frac{D}{W_e}\right) = \log_2\left(1+\frac{D}{4.133\sigma}\right) \qquad \text{bit.} \qquad (2)$$

Historically, Fitts' law emerged in a specific context: Information Theory was booming[2], and at the time many psychologists (e.g., [1, 9, 14]) were trying to apply its operational results to human behaviour. One concept crucial to most researchers, Fitts included, was the *channel capacity*, namely, the maximum rate at which information can be reliably transmitted in a communication system. The expression for ID shown in Equation (1) is based on Shannon's formula for the capacity of the Gaussian channel [3, 4, 12].

Errors vs. Erasures
In this paper, we call attention to the distinction made in Information Theory between two sorts of transmission mistakes: *errors* vs. *erasures*. We will show that this classic distinction, which apparently has escaped the attention of Fitts' law students, opens the way to an elegant theoretical solution to

[2] Shannon himself believed it was even ballooning [16].

the problem raised in the Fitts paradigm by the non-constant frequency of target misses.

In communication engineering, an *error* is said to have occurred when the received symbol differs from that originally sent. For example, the word BUTTER is received in the place of the sent word BATTER, the A having been accidentally replaced by an U. But suppose that the received word is B?TTER, with the question mark signaling a missing character: This is what is called an *erasure*. One important difference between an error and an erasure is that the former conveys wrong information whereas the latter conveys no information at all.

How does the error/erasure distinction apply to Fitts' target-reaching paradigm? Below we will argue that the option most consistent with the logic of this paradigm is to model target misses as erasures, rather than errors.

MEASURING ACCURACY IN FITTS' PARADIGM

The goal of a Fitts' law experiment being to observe and study the speed-accuracy tradeoff described by Woodworth [20], the choice of the metrics used to measure *speed* and *accuracy* is fundamental. While there has been unanimous agreement in the literature that movement time provides a satisfactory measure of speed [8], the measure for accuracy defined by Equation (1) has been controversial from the outset [3]. It is only recently that the adjustment for errors was standardized by ISO [10, 17].

In a Fitts' law experimental setup[3], the task of the participant is to aim for a target of predefined width as fast as possible while making as few mistakes (target misses) as possible. There are three different ways of handling these mistakes: simply ignoring them, measuring the error rate (a percentage), or measuring the spread of endpoints (typically a standard deviation).

- *Ignoring the mistakes*. Fitts, who did not measure actual amplitudes, classified the movements in an all-or-none way as hits and misses. Although he did tabulate the (variable) error rates he obtained in his experiments, he felt in a position to leave them aside because of the "small incidence" of target misses [4, p. 265].

- *Taking the error rate into account*. To our knowledge, Crossman [3] was the first to try to incorporate the error rate information into his *ID* measure, leveraging the standard Gaussian distribution model.

- *Taking the spread of endpoints into account*. This is the standardized way of measuring accuracy in Fitts' law [10, 17]. Recourse to the standard deviation as a measure of accuracy has the implication that the amplitude of the mistake matters in the upcoming analysis: if the target is missed, the farther from the edge of the interval, the worse the performance. It also implies that there is equivalence between two movements hitting the target if and only if they end up at exactly the same distance from the center of the target.

[3] While Fitts's time-minimization paradigm is widespread in HCI, it is important to bear in mind that it is not the only workable paradigm for studying the speed-accuracy trade-off of aimed movement [7, 8, 15].

The ISO standard and Fitts' law literature in general treats mistakes as *errors*, by referring to the standard deviation of the endpoints distribution—either by direct estimation or through a calculation from error rates. Thus in the *error* concept, the accuracy depends on the (continuous) distance between the movement endpoint and the target center.

There seems to be a contradiction between this approach and the all-or-none logic of Fitts' experimental paradigm: Since the only instruction for participants is to *hit* the target about 96% of the time (i.e., *miss* it about 4% of the time), all movements that end up inside the W interval should be considered equivalent, in keeping with the observation that in a real-world interface, what matters is not precisely *where* the click takes place, but rather *whether or not* the click falls in the intended area. This corresponds to the information-theoretic concept of *erasures*.

Besides this conceptual mismatch between errors and mistakes, ID_e suffers from at least three further deficiencies.

INFORMATION-THEORETIC CRITIQUE OF ID_e

The detailed expression of the effective width W_e in Equation (2) as given in [17] is as follows. Let σ denote the standard deviation of the end-point distribution, and ε the error rate, i.e., the proportion of target misses:

- If σ is available:

$$W_e = 4.133\sigma. \tag{3}$$

- Otherwise:

$$W_e = \begin{cases} W \cdot \frac{2.066}{z(1-\varepsilon/2)} & \text{if } \varepsilon > 0.0049\% \\ 0.5089 \cdot W & \text{otherwise.} \end{cases} \tag{4}$$

The received justification is as follows [12, Section 2]:

> "The entropy (H), or information, in a normal distribution is $H = \log_2\left((2\pi e)^{1/2}\sigma\right) = \log_2(4.133\sigma)$, where σ is the standard deviation in the unit of measurement. Splitting the constant 4.133 into a pair of z-scores for the unit-normal curve (i.e., σ = 1), we find that the area bounded by z = $\pm$2.066 represents about 96 % of the total area of the distribution. In other words, a condition that target width is analogous to the information-theoretic concept of noise is that 96 % of the hits are within the target and 4 % of the hits miss the target [...]. When an error rate other than 4% is observed, target width should be adjusted to form the effective target width in keeping with the underlying theory."

We see three issues with ID_e:

1. The computation of W_e as 4.133σ as well as the computation leading to Equation (4) presumes a Gaussian distribution of endpoints [17], but the validity of this hypothesis has been questioned empirically, e.g., [4] [19].

2. To our knowledge Information Theory provides no justification to the relation $W_e = 4.133\sigma$. When Crossman [3] calculated the expression for W_e from the area under the standard normal curve, he took the 5% value as an arbitrary "permissible" error rate. McKenzie [12] noticed that

by changing the arbitrary rate from 5% to 3.88% (approximately 4%), the entropy of the rectangular distribution of width W_e would equal the entropy of the Gaussian distribution of standard deviation σ (see Appendix), but this is no more than a nice coincidence: we can see no information-theoretic reason to equalize these two entropies.[4].

3. The threshold of error rate placed at 0.0049% is arbitrary. Even if considering a Gaussian distribution for the endpoints, the one-to-one relationship between standard deviations and error rates is only true for strictly positive error rates. Indeed, when the error rate is vanishing, the standard deviation is also vanishing, so ID_e tends to infinity! To prevent this from happening, [17] recommends that below a certain error rate (0.0049%), ID_e should be kept constant. The justification of the threshold error rate of 0.0049% is that it "rounds to 0.00". As shown below, the existence of such a threshold is in fact adverse to the theory.

The standardized index of difficulty ID_e is thus questionable. It relies on the unsafe Gaussian hypothesis, two arbitrary constants, and one coincidence. Even more importantly, it has never been shown to be the correct expression of the capacity of a human-motor channel—the expected rationale behind Fitts' law.

We now propose a new effective index $ID(\varepsilon)$ that is compliant with Fitts' experimental design, does not need the Gaussian hypothesis and is justified theoretically as a channel capacity.

A COMPLIANT INDEX OF DIFFICULTY: $ID(\varepsilon)$

We have already noted that treating target misses as errors is not adapted to Fitts' paradigm—these events should rather be viewed as erasures. Indeed, the design of the experiment entails a binary decision: there is a target and the movement either finishes inside (a hit) or outside (a miss). This is consistent with the instruction "try to hit the target" as opposed to "try to aim for the center of the target". In this section we build a model for a channel which allows erasures, and obtain a new index of difficulty $ID(\varepsilon)$ through Shannon's concept of channel capacity.

Consider a channel that oscillates randomly between a good (G) state and a bad (B) state, with probability ε of being in state B and probability $1-\varepsilon$ of being in state G. When the channel is in its good state, it corresponds to the typical channel of capacity $\log_2\left(1+\frac{D}{W}\right)$ described by McKenzie [12], which we refer to as the *Fitts channel*. However, when the channel is in its bad state it can only produce erasures—we call it an *erasure channel*. This configuration (Figure 1) is known as a *compound* channel [5].

We now evaluate Shannon's *capacity* of such a channel as a common ground to compare the performance of different

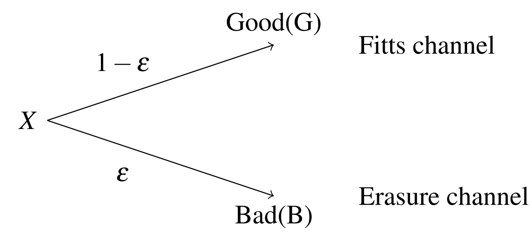

Figure 1. Compound channel for an aiming task with target misses.

participants operating at different accuracy levels (with different values of ε). The channel capacity corresponds to the maximum transmission rate that the participants would have achieved with an arbitrarily small error rate, as if they never missed the target. Shannon's capacity of the compound channel of Figure 1 is given by the following theorem.

THEOREM (COMPOUND CHANNEL CAPACITY). *Consider a compound channel as in Figure 1, with probability ε of being in state B and probability $1-\varepsilon$ of being in state G. The capacity of such a channel is given by*

$$C = (1-\varepsilon)\log_2\left(1+\frac{D}{W}\right).$$

As expected, the obtained capacity is lower than the capacity $\log_2\left(1+\frac{D}{W}\right)$ that would have been achieved with 100% hitting accuracy ($\varepsilon = 0$).

The formal information-theoretic proof is known [2] and summarized in the Appendix for completeness, but is easy to sketch in order to understand the reasoning: The participant is effectively time sharing both channels. With Fitts' channel, the transmitted information is $\log_2\left(1+\frac{D}{W}\right)$ bits and with the erasure channel the transmitted information is 0 bit, so that, on average, $C = (1-\varepsilon)\times\log_2\left(1+\frac{D}{W}\right)+\varepsilon\times 0$. In line with Fitts' parallel between capacity and ID, our new effective index is

$$ID(\varepsilon) = (1-\varepsilon)\log_2\left(1+\frac{D}{W}\right)$$

where ε is no other than the traditional 'error rate' (more cautiously designated here as the percentage of target misses).

COMPARING THE TWO INDICES

We now provide an analytical comparison of ID_e and $ID(\varepsilon)$. The behavior of the standardized ID_e for vanishing error rates is problematic. Indeed, the inverse Gauss error function (see Appendix) $\mathrm{erf}^{-1}(1-\varepsilon)$ tends to $+\infty$ as ε vanishes, so that we would normally have

$$\lim_{\varepsilon\to 0} ID_e = \infty.$$

Due to the 0.0049% bounding, however, instead we obtain

$$\lim_{\varepsilon\to 0} ID_e = \log_2\left(1+\frac{D}{0.5089W}\right) \simeq \log_2\left(1+\frac{2D}{W}\right) = 1+\log_2\left(\frac{1}{2}+\frac{D}{W}\right),$$

[4]Incidentally, these entropies can both be negative. Information Theory distinguishes the (discrete) entropy of a discrete random variable, which is non-negative and serves as a measure of information, and the (so-called *differential*) entropy of a continuous random variable such as a normal random variable, which is positive for large variances and negative for small variances and thus cannot be interpreted as a measure of information [2]

which is equivalent to the Welford index of difficulty [18]. The particular choice to bound the index at the 0.0049% rate results in the index coincidentally tending to the Welford ID, not the Shannon ID. Therefore, there is no continuity as epsilon approaches zero for ID_e.

In contrast, $ID(\varepsilon)$ *does* have the property of continuity towards zero since obviously $ID(0) = ID$.

Figure 2 shows the two indices $ID(\varepsilon)$, ID_e as well as the unbounded $u\text{-}ID_e$ (for which the 0.0049% distinction is not made) for $D/W = 15$ as a function of ε in the interval $[0-1]$. The difference $ID_e - ID(\varepsilon)$ between ID_e and $ID(\varepsilon)$ is lowest around $\varepsilon = 0.1$. With higher values of ε, the difference increases but such high errors rates are not common. However, for very small values of ε, $ID(\varepsilon)$ can be up to 1 bit smaller then ID_e. It can be noted that this difference accounts for very careful participants which are well handled by $ID(\varepsilon)$.

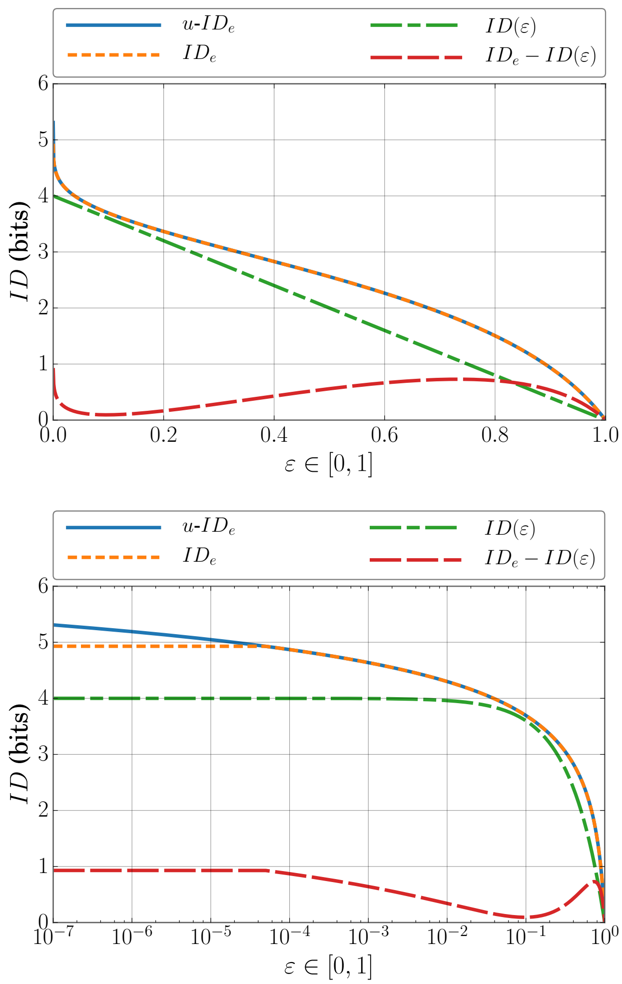

Figure 2. Comparison of $ID(\varepsilon)$ and ID_e for erasure rate in $[0,1]$, for $\frac{D}{W} = 15$. $u\text{-}ID_e$ refers to ID_e where the 0.0049% distinction is not made. The scale is lin-lin above and lin-log below.

IMPLICATIONS FOR HCI

As we have seen, there remains room in Fitts' law research to improve the theory. We feel that the erasure-channel notion has promise to improve our conceptual grasp of accuracy, the Achilles heel of Fitts' experimental paradigm.

Capitalizing on the fact—so far unrecognized, to our knowledge—that the target misses that take place in Fitts' time-minimization paradigm can be interpreted as erasures in the language of Information Theory, we have derived a new effective index of difficulty which has a number of desirable properties: not only is it theoretically safer than the ISO index as it does not presuppose a Gaussian distribution of endpoints, it is also theoretically more relevant, having been proven to measure a capacity in Shannon's strict sense. It is also more convenient in practice for two reasons: (a) it is continuous towards the zero-error rate region, thus allowing the researcher to dispense with an arbitrary treatment of the 0 percent miss case; and (b) it is a great deal simpler to compute than the ID_e traditionally used in HCI by Fitts' law practitioners.

This is a purely theoretical note and obviously more research is needed to seriously evaluate all the implications of our result. The reader might feel that perhaps the above-described index is just a better tool to obtain essentially the same results, but there is nothing so practical as a good theory, and there is reason to think that a conceptually neater and practically easier ID should help the HCI researchers who routinely exploit Fitts' law to obtain safer results.

ACKNOWLEDGMENTS

The authors would like to thank Michel Beaudouin-Lafon for his precious help during the preparation of the manuscript.

APPENDIX

PROOF FOR W_e. Consider the endpoint location Y about the target center, assumed normally distributed $\mathcal{N}(0,\sigma^2)$. The event $|Y| > \frac{W}{2}$ defines a target miss. Then target width W and error rate ε are related by:

$$\varepsilon = 1 - \operatorname{erf}\left(\frac{W}{2\sqrt{2}\sigma}\right) \tag{5}$$

which allows to link W and W_e. Equalizing $\log_2 W_e$ to the Gaussian entropy $H = \log_2(\sqrt{2\pi e}\sigma)$ gives $W_e = 4.133\sigma$, which from Equation (5) amounts to choose $\varepsilon = 3.88\%$. □

PROOF OF THE THEOREM. The channel capacity [2] is defined as the maximized value of mutual information $I(X;Y)$. Let S ($= G$ or B) be the channel state. Since observing the output Y already determines the state, one has $I(X;Y) = I(X;(Y,S))$. By the chain rule for mutual information [2], one has

$$\begin{aligned} I(X;(Y,S)) &= \mathbb{P}(S=G)\cdot I(X;Y|S=G) + \mathbb{P}(S=B)\cdot I(X;Y|S=B) \\ &= (1-\varepsilon)\cdot I(X;Y|S=G) + \varepsilon\cdot I(X;Y|S=B). \end{aligned}$$

Here $I(X;Y|S=G)$ is the mutual information computed for the Fitts channel, and $I(X;Y|S=B) = 0$ bit since for bad state, only an erasure can be output. Thus $I(X;Y) = (1-\varepsilon)I(X;Y|S=G)$. Maximizing over the input distribution gives the result. □

REFERENCES

1. F. Attneave. 1959. *Applications of information theory to psychology: A summary of basic concepts, methods, and results.* Henry Holt.
2. T. M. Cover and J. A. Thomas. 2012. *Elements of information theory*. John Wiley & Sons.
3. E. R. F. W. Crossman. 1957. The speed and accuracy of simple hand movements. *The nature and acquisition of industrial skills* (1957).
4. P. M. Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.
5. E. N. Gilbert. 1960. Capacity of a Burst-Noise Channel. *Bell system technical journal* 39, 5 (1960), 1253–1265.
6. D. M. Green and J. A. Swets. 1966. Signal detection theory and psychophysics. 1966. *New York* 888 (1966), 889.
7. Y. Guiard, H. B. Olafsdottir, and S. T. Perrault. 2011. Fitt's law as an explicit time/error trade-off. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1619–1628.
8. Y. Guiard and O. Rioul. 2015. A mathematical description of the speed/accuracy trade-off of aimed movement. In *Proceedings of the 2015 British HCI Conference*. ACM, 91–100.
9. W. E. Hick. 1952. On the rate of gain of information. *Quarterly Journal of Experimental Psychology* 4, 1 (1952), 11–26. DOI: http://dx.doi.org/10.1080/17470215208416600
10. ISO9241. 2000. 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs)-Part 9: Requirements for non-keyboard input devices (FDIS-Final Draft International Standard), 2000. *International Organization for Standardization* (2000).
11. I. S. McKenzie. 1989. A note on the information-theoretic basis for Fitts' law. *Journal of motor behavior* 21, 3 (1989), 323–330.
12. I. S. McKenzie. 1992. *Fitts' law as a performance model in human-computer interaction*. Ph.D. Dissertation. University of Toronto.
13. I. S. McKenzie. 2002. Bibliography of Fitts' law research. (2002). http://www.yorku.ca/mack/RN-Fitts_bib.htm [Online; last update: June 25, 2002].
14. G. A. Miller. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63, 2 (1956), 81–97.
15. Richard A Schmidt, Howard Zelaznik, Brian Hawkins, James S Frank, and John T Quinn Jr. 1979. Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychological review* 86, 5 (1979), 415.
16. C. E. Shannon. 1956. The bandwagon. *IRE Transactions on Information Theory* 2, 1 (1956), 3.
17. R. W. Soukoreff and I. S. McKenzie. 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 (Dec. 2004), 751–789. DOI: http://dx.doi.org/10.1016/j.ijhcs.2004.09.001
18. A. T. Welford. 1960. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics* 3, 3 (1960), 189–230.
19. A. T. Welford, A. H. Norris, and N. W. Shock. 1969. Speed and accuracy of movement and their changes with age. *Acta psychologica* 30 (1969), 3–15.
20. R. S. Woodworth. 1899. Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements* 3, 3 (1899), i.