RJ 10253 (A0208-026) August 21, 2002
Computer Science

IBM Research Report

On the Validity of Throughput as a Characteristic of Computer Input

Shumin Zhai
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

LIMITED DISTRIBUTION NOTICE: This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. Ithas been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distributionoutside of IBM prior to publication shoukl be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g, payment of royalties). Copies may be requested from IBM T. J. Watson Research Center , P. O. Box 218, Yorktown Heights, NY 10598 USA (email: reports@us.ibm.com). Some reports are available on the internet at http://domino.watson.ibm.com/library/CyberDig.nsf/home

(Working paper, comments welcome)

On the Validity of Throughput as a Characteristic of Computer Input

Shumin Zhai

IBM Almaden Research Center

650 Harry Road, San Jose, California

zhai@almaden.ibm.com

Abstract

Throughput (*TP*) has been a fundamental metric in quantifying input system performance. *TP* is a concept based on Fitts' law, an essentially two parameter (*a, b*) relationship between movement time (*MT*) and Fitts' index of difficulty (*ID*). In part thanks to ISO 9241-9, the final draft international standard (FDIS) of "Ergonomic requirements for office work with visual display terminals - Part 9: Requirements for non-keyboard input devices", research and testing of computer input system in recent years have increasingly relied on *TP* as the sole measure of performance quality of input devices. The goal was to standardize studies onto one metric that can be generalized and compared across different experimental studies. Unfortunately *TP* as defined in ISO 9241-9 is an ill-defined concept that changes with the mean index of difficulty used in measuring it and therefore cannot be generalized beyond specific experimental settings. Furthermore, important properties can be hidden when *TP* is used as a single metric to input system evaluation. We reason that it is more informative to use (*a*, *b*) parameters in Fitts' law as separated metrics of an input system. One related issue, the foundation for post-hoc target size adjustment, known as effective width, is also discussed.

Categories and subject descriptors: H.5.2 User Interfaces: *Benchmarking, Evaluation/methodology, Graphical user interfaces, Input devices and strategies, Standardization, Theory and method.* H.1.2 Models and Principles: *User/Machine systems – Human information processing*

General terms: Human factors, Performance, Theory

Key words and phrases: computer input, Fitts' law, motor control performance, throughput, bandwidth, index of performance (*IP*), ISO 9241-9.

Introduction

One important goal of computer input research is to measure and characterize the usability and performance of various input systems[1]. Naturally, the more generalizable the characterization is, the more useful it is. It is difficult to make use of performance characterizations that cannot be generalized. For example, if an input device A (e.g. a mouse) were said to be superior to an input device B (e.g. a trackball) for pointing tasks, we would have to ask by what metrics such a statement is made. If we say system A was measured to take 0.9 seconds and system B 1.0 seconds on average to perform a pointing action, with similar error rate (e.g. 4%), a carefully recipient of that information would have to ask under what

[1] For brevity, we use the term input system to mean one, a few, or all aspects of computer input – device, sensor, transfer function, interaction technique, user expertise, shape and size the device handle, etc.

conditions such measurements were made. If the measurements were made with an icon of one centimeter in width, over 10 centimeters distance, the recipient may not find that information useful because his or her application might be primarily concerns with much smaller targets.

Card, English and Burr first realized that human performance models could be used to generalize studies on input devices (Card, English, & Burr, 1978). In particular, they applied the well-known Fitts' law to computer pointing tasks:

$$MT = a + bID \qquad (1)$$

where *MT* is movement time and $ID = log_2 (2D/W)$ is the index of difficulty as originally defined by Fitts (Fitts, 1954). *D* and *W* are target distance and size respectively, and *a* and *b* are empirically determined constants, depending on the input system used. An obvious problem with this definition of *ID* is that when *D*, the distance from the center of one target to another, is zero, *ID* (and hence *MT*) tends to negative infinity. MacKenzie (MacKenzie, 1989) first proposed to return to Shannon's original formula of information in a noisy channel, which was the inspiration of Fitts' law to define index of difficulty, to quantify *ID*, i.e.: $ID = \log_2 (D/W +1)$, which resolves the negative *ID* problem.

Important to the use of Fitts' law in this context is that the infinite number and range of possible target sizes and distances can be unified into one variable *ID*. Quantifying input systems hence can be made by measuring *a* and *b* parameters in Fitts' law. If done properly, such metrics would be independent of the experimental task setting and generalizable to other target sizes and distances[2].

In recent years, in part thanks to ISO 9241-9, the final draft international standard (FDIS) of "Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices", research and testing of computer input system have increasingly relied on throughput ($TP = ID / MT$) that further combines *a* and *b* into one quantification metric. Research on a number of devices and techniques have been conducted in such a manner, ranging from isometric joystick ((Douglas, Kirkpatrick, & MacKenzie, 1998)), touchpad (Douglas et al., 1998), remote control mouse (MacKenzie & Jusoh, 2001), (Silfverberg, MacKenzie, & Kauppinen, 2001) and laser pointers (Myers et al., 2002) (Oh & Stuerzlinger, 2002). The goal was to standardize studies onto one metric that can be generalized and compared across different experimental studies.

What has not been realized by the authors of ISO 9241-9 and these studies is a simple fact that in general *TP* is not a constant independent of the mean index of difficulty used in measuring it hence cannot be generalized beyond the task parameters used in a study. In other words, there lack a logical foundation to the validity of *TP* as an intrinsic performance characteristic. Furthermore, important properties can be hidden when *TP* is used as a single metric to input system evaluation. There are two aspects to *TP* as defined in ISO 9241-9. One is incorporating error into movement time metric by using effective rather than nominal target width, the other is reducing two parameters of Fitts' law (*a*, *b*) into one metric by the division of mean *ID* and mean *MT*. We analyze these two aspects separately in the rest of the paper.

Incorporating error into movement time

The smaller number of metrics needed to characterize an input system, the more convenient and powerful they are. Other than *a* and *b* that quantifies the time aspects of Fitts' law, participants may also make errors during pointing experiments. Ideally the experimental participants follow the instruction of performing pointing "as quickly possible and as accurately as possible" and only miss the target in about 4% of the trials as a result. However in reality error rate may vary across systems, *ID*'s or studies. This means that

---

[2] This is only true within a reasonable range, but not to extremely large or small scale ends (cf. (Guiard, 2001), (Accot & Zhai, 2001)).

one cannot use *MT* alone to quantify the performance of an input system. Sometimes *MT* and error rate change in the same direction from one input system to another and other times in opposite directions, making it difficult to conclude which system is overall superior.

A technique adopted by ISO 9241-9 FDIS and used by some (e.g. (Welford, 1968) (MacKenzie, 1992), (Douglas et al., 1998; MacKenzie & Jusoh, 2001) (Myers et al., 2002; Silfverberg et al., 2001) (Oh & Stuerzlinger, 2002)) but not all researchers (e.g. (Fitts, 1954), (Fitts & Peterson, 1964), (Card et al., 1978), (Mottet, Guiard, Ferrand, & Bootsma, 2001), (McGuffin & Balakrishnan, 2002), (Accot & Zhai, 2002)), is to incorporate error into *MT*. This was done by post hoc adjustment of nominal width *W* to effective width W_e in calculating *ID*. W_e was so chosen that effectively 4% of the trials fell outside the adjusted target[3]. The underlining assumption is that if pointing movement is indeed like information transmission as hypothesized by Fitts, then the hits (the landing points of pointing) should be distributed along a Gaussian distribution curve. Furthermore errors can be traded off with *ID* (hence *MT*) along the Gaussian distribution curve (Figure 1). This is to say, for example, making 14% error with *MT* time at a target of *W* width is equivalent to making 4% error at a target of W_e width with $MT_e = a + b\log_2(D/W_e + 1)$ time, where W_e is chosen as if 10% of the error falls inside the target and 4% outside the target.

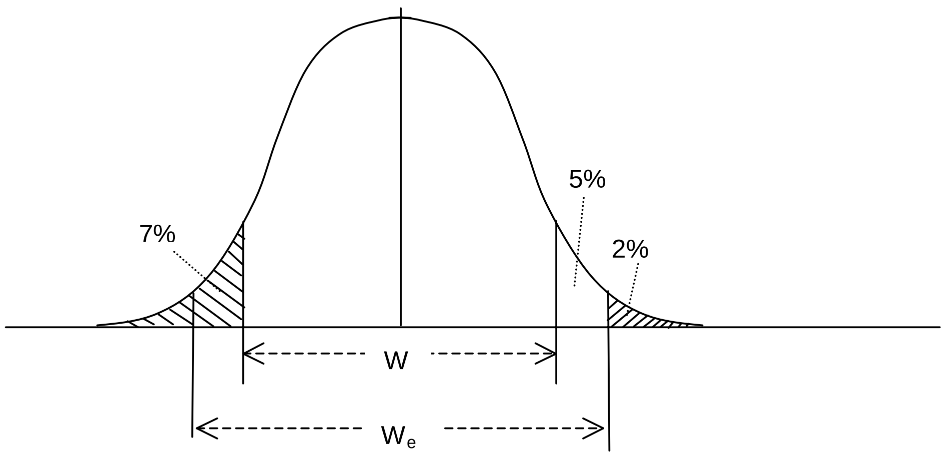

Figure 1 . An illustrative example: adjustment from nominal width *W* to "effective" width W_e so 4% (2% on each side) of trials fall outside W_e.

There could be three lines of arguments or support for the validity of such a conversion – a matter of standardization, empirical evidence, and theoretical plausibility.

The standardization argument could be that this is an agreement on how error should be converted into time, with or without basis. Once agreed, all research should be done in such a manner so there is a consistent method. A similar standardization method to speed and accuracy trade-off is in typing contest, where each stroke of mistake is counted as five words less in the final score. The counter argument to this conversion agreement argument is that it may or may not be "fair" to specific applications. In some

---

[3] W_e can be either calculated from standard division of the hits or converted from actual error rate based on Gaussian distribution z-score (MacKenzie, 1992). Both methods assume Gaussian distribution and the effect is the same: leaving 4% of trials outside of adjusted W_e.

applications, error is much more critical than others. How error is weighed against time should be a decision made by the users of input performance information based on their application.

An empirical argument could be an experiment where target width and error rate are systematically manipulated, which allows quantitative examination of error and target width relationship and see if they indeed trade along the Gaussian distribution curve. Unfortunately no such a study has been found in the literature. We call for such a study in the future.

A theoretically argument could be that Fitts' law tasks follow information theory and has to behave like information transmission in a noisy channel. More noise has to reduce information in a quantitative way that supports the W_e conversion. While Fitts' law as an empirical relationship is widely confirmed, its theoretical basis is not agreed in the literature at all. Fitts' original theoretical foundation was information theory – the mathematical theory of communication ((Fitts, 1954), (Shannon and Weaver, 1949)). He viewed the motor control system quite literally as a communication channel with noise, based on the argument that subjects in fact use all the tolerance specified with an approximate Gaussian distribution, exhibiting exactly the amount of information specified by that task, no more (wider than target size) and no less (narrower than the target size). Interestingly, according to such reasoning the spread of hits should not be much greater or much less than the nominal target width. Otherwise the task does not behave like information transmission along Fitts' argument.

Fitts' information transmission theory on his lawful relationship has been challenged by other researchers. For example Crossman and Goodeve (Crossman & Goodeve, 1963/1983) argued that a feedback control theory was more appropriate. They first proposed a theory in which the effector was moved at a speed negatively proportional to the remaining error (first order velocity control). However they rejected this theory upon further examination and opted for an ``intermittent sampling and proportional correction model'' which postulated reaching movement with a series of (sub)movement corrections, each with an amplitude proportional to the remaining error to the target. Schmidt and colleagues (Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979) noted the theoretical difficulty in such a theory because it involved four or five corrections in one trial of reaching, each requiring at least 120 to 200 ms response time. Instead, they attempted to explain the speed-accuracy tradeoff by motor-output variability (Schmidt et al., 1979). An interesting result of Schmidt and colleagues' endeavor is Schmidt's law that is a linear relationship between movement time and the ratio of target distance and its "effective width" – defined by the standard deviation of the hits – in reaching tasks. A critical difference between Schmidt's law and Fitts' law lies in the role of dependent and independent variables in human performance. Fitts' law treats spatial properties of movement (D and W) as independent variable (controlled and manipulated in experiments) and temporal performance as dependent variable (measured from the experiments). In contrast, Schmidt's law specifies temporal property (the time or pace of the reaching task) and measures the spatial property of movement (the effective width of the reaching movement). Fitts' law applies to close-loop information-processing based reaching tasks, while Schmidt's law applies to open-loop reaching task dictated by motor impulse variability (Schmidt et al., 1979). Along this line of thinking, when participants did not use the nominal width (independent variable) specified in Fitts' law task, the task starts to move away from Fitts' paradigm. Therefore post-hoc adjustment to independent variable is on a slippery ground.

To complete this short review of Fitts' law theories, Meyer and colleagues (Meyer, Abrams, Kornblum, Wright, & Smith, 1988) proposed a theory that advances the Crossman-Goodeve deterministic model of iterative-correction with stochastic components. Their theory divides reaching movement into two sub-movements and views the total movement as a result of optimization or ``ideal compromise'' between the durations of primary and secondary sub-movements. If and how such a theory can support effective width adjustment is an open question.

Another unresolved issue in converting error to W_e is if such a conversion should be done on the basis of an individual or the whole group of experimental participant. It is more theoretically plausible to use trials from individual participant on one *W* and *D* combination, as these are from the one source of "information channel". A side effect of this approach is that ID_e adjusted in such a way may vary from one participant to another, resulting different points for Fitts' law regression analysis. This weakens the averaging effect that eliminates individual difference and other "noises" in Fitts' law regression. In addition the degree of freedom in regression increases *n* folds (*n* is the number of participants). The goodness of fit in Fitts law regression will hence suffer accordingly.

Some studies, such as the re-analysis of Fitts' data based on W_e in (MacKenzie, 1992), calculate W_e on a group basis. Given the trials assembled in such a manner are not from the same information channel, its theoretical foundation is sound than adjustment on individual basis.

To conclude, while the intent of using W_e to incorporate error into *MT* is compelling for pointing tasks[4] and may indeed be proven valid in the future, we have not found enough theoretical or empirical bases for such a conversion in the literature. Before then, it is recommendable to additionally report error rate and *MT* parameters (*a, b*) based on unadjusted *W* separately, if the W_e approach is used.

Reducing a and b into one metric TP

Throughput (*TP*), a metric based on Fitts' law with or without using W_e conversion, has also been called index of performance (*IP*) or bandwidth. Since Fitts' law can quantify pointing task in terms of information with *ID* (in bits), conversely the amount information transmitted (or more accurately expressed) per unit time (in bits per second) is a logical metric to quantify the performance of an input system.

There have been at least two ways to calculate *TP*. One is to use the inverse of *b* in Equation 1. For clarity, we used TP_b to denote this approach:

$$TP_b = 1/b \qquad (2)$$

b is in the unit of second/bits; the unit of TP_b is hence bits per second (bps, sometimes also denoted bits/s or b/s).

This method, used by many researchers (e.g. (Card et al., 1978), (MacKenzie, Sellen, & Buxton, 1991), (Soukoreff & MacKenzie, 1995), (Zhai, Morimoto, & Ihde, 1999), (Zhai, Smith, & Hunter, 2002a)), does not incorporate parameter *a*, the Fitts' law regression intercept, and hence is only one part of movement time. It has to be considered in conjunction with *a*. Ignoring constant *a* can be misleading. For example, using $TP_b = 4.9bps$ and considering *a* = 0 except when tapping on the same key consecutively, Zhai, Hunter and Smith (Zhai, Hunter, & Smith, 2000) (Zhai et al., 2002a) estimated movement efficiency of various stylus keyboard layouts. Later our research based on empirical data showed that *a* could not be neglected and hence the performance estimations had to be modified (Zhai, Sue, & Accot, 2002b).

The second method, used by researchers including Fitts' himself (Fitts, 1954) and increasingly by many others thanks to its adoption in ISO 9241-9 FDIS (Douglas et al., 1998; MacKenzie & Jusoh, 2001) (Myers et al., 2002; Silfverberg et al., 2001) (Oh & Stuerzlinger, 2002)) incorporates *a* into *TP* calculation. We use TP_a to denote this calculation:

---

[4] ISO 9241-9 FDIS also recommends using W_e to calculate path steering task difficulty, defined as *ID* = *D/W* where *D* is the distance and *W* is the width of a linear path (Accot & Zhai, 1997). Such a conversion in the case of steering has no plausible basis since the *law of steering* only models successful steering trials. The nature of error in steering tasks, which may occur anywhere along the steering path, is also fundamentally different Fitts' task where error only occurs on the final landing point.

$$TP_a = \overline{ID} \,/\, \overline{MT} \,, \qquad (3)$$

$\overline{ID}$ and $\overline{MT}$ are mean values of *ID* and *MT* in an experiment. *ID* may or may not be adjusted for error using W_e although ISO 9241-9 FDIS suggests so.

When $a = 0$ (a special case), the two approaches give the same result:

$$TP_a = \overline{ID} \,/\, \overline{MT} = \overline{ID} \,/\, \overline{(a + bID)} = 1/b = TP_b \,, \qquad (4)$$

The goal and a potential advantage of using TP_a is to characterize pointing performance of an input system into one value. Such a value should be independent of, or at least generalizable beyond, the task parameters such as the number, sizes, and distances of targets used in measuring *TP*. If so, TP_a would be intrinsic to an input system. However this is not true when *a* is not zero.

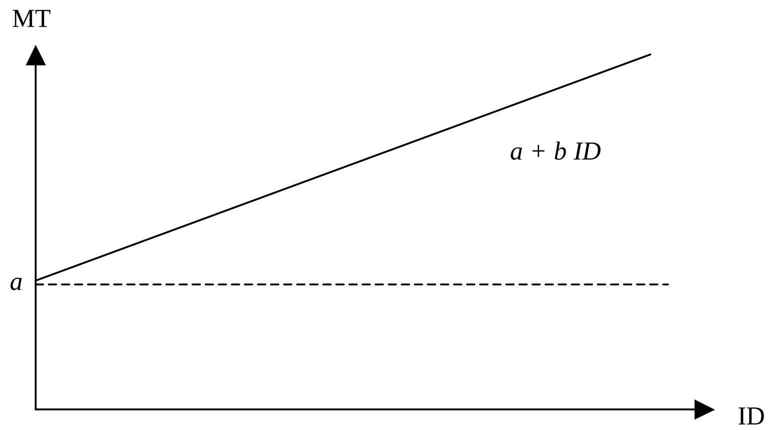

Figure 2. Fitts' law: movement time in pointing increases linearly with *ID*, in addition to the constant *a*.

According to Fitts' law ($MT = a + bID$, Figure 2) pointing time cannot be entirely quantified by information terms (namely *ID*) when *a* is not zero. There is a constant portion (*a*) in completion time regardless the amount of information expressed. $TP_a = \overline{ID} \,/\, \overline{MT}$ incorporates both *a* and *b* into one metric. On the surface, this seemed have solved the problem of a non-zero *a*, but in fact it means TP_a is no longer an intrinsic value to an input system, but rather is a value dependent on the task parameters.

An input system test based on Fitts' law is usually carried out with systematic manipulation of *ID* through target size and distance (*W*, *D*) combinations. Suppose the number of *ID* in an experiment is *N*, then according to (4) TP_a is calculated as

$$TP_a = \frac{\sum_{i=1}^{N} ID_i}{N} \Bigg/ \frac{\sum_{i=1}^{N} MT_i}{N} \qquad (5)$$

Since *MT* follows Fitts' law, we may substitute *MT* with *a* + *b ID*, and obtain:

$$TP_a = \frac{\sum_{i=1}^{N} ID_i}{N} \bigg/ \frac{\sum_{i=1}^{N} (a + bID_i)}{N} = \overline{ID} / (a + b\overline{ID}) \qquad (6)$$

Clearly, TP_a depends not only on *a* and *b*, two intrinsic quantities to the input system, but also the mean *ID* used in the experiment. Figure 2 illustrates how TP_a changes with mean *ID* with typical Fitts' law parameter values *b* = 0.125 (sec/bit) and *a* = 0.1, 0.2 or 0.4 sec.

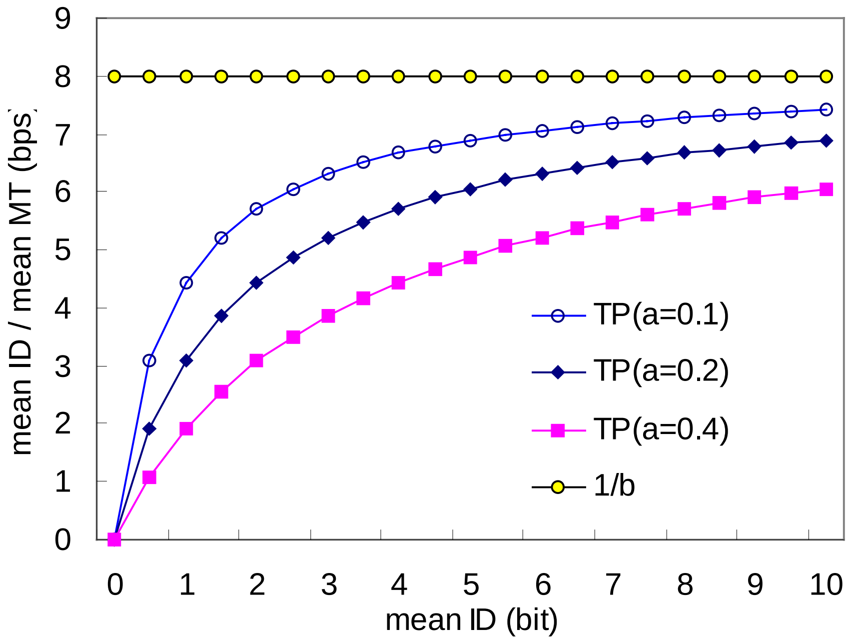

Figure 3. Throughput defined by $TP_a = \overline{ID} / \overline{MT}$ as a function of mean ID.

In sum, TP_a is not a system intrinsic constant. When a throughput value is reported for an input system, its meaning is ill-defined, because there is a free degree of freedom (mean *ID*). Depending on mean *ID*, TP_a moves up and down on the curves in Figure 2. In other words, TP_a is not a metric that can be generalized beyond experimental parameters.

The sensitivity of TP_a to the mean *ID* used varies with *a*. The greater *a* is, the more slowly TP_a converges to 1/*b*. When *a* = 0, $TP_a = TP_b$ = 1/b.

TP_a has an asymptote of *1/b* when mean *ID* increases toward infinity. This means that a more stable value of TP_a can be measured if a greater mean *ID* is used in the experiment. On the other hand this is totally unnecessary because the asymptote can be easily estimated from Fitts' law regression (1/*b*).

Sometimes a negative *a* value is reported. When *a* is negative, TP_a calculation is even more unstable. Not only TP_a still is not a constant. It in fact has a singular point at $ID = |a|$ where TP_a is infinity. Figure 3 shows TP_a as a function of mean *ID* when *a* = - 0.15 sec.

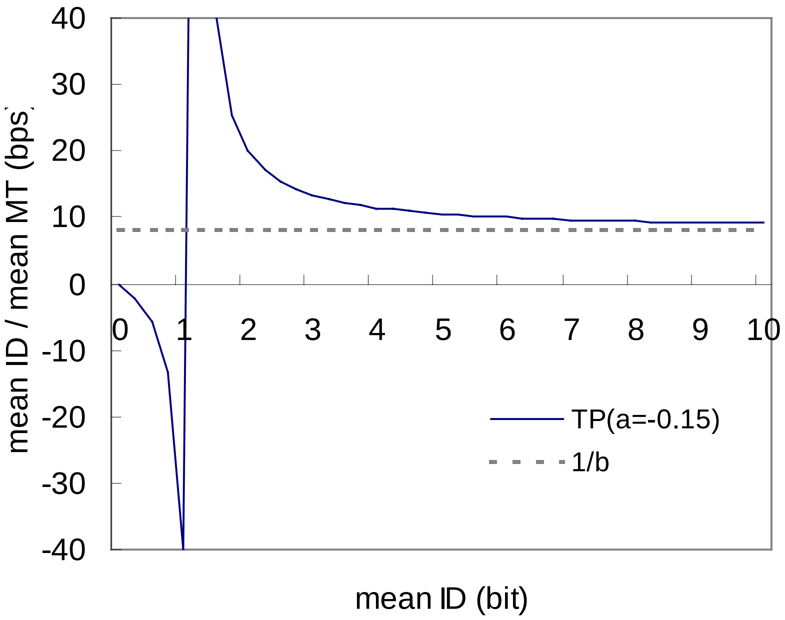

Figure 4. $TP_a = \overline{ID} / \overline{MT}$ **as a function of mean ID, when *a* is negative.**

Now that the dependency of TP_a on mean *ID* is understood, is it possible to standardize mean *ID* used in experiments so TP_a can be compared across studies? This is difficult because the range of *ID* in an experiment depends on the practical application that motivates the study. In practice even similar studies by the same researchers could change the set of *IDs* in different studies. For example, MacKenzie & Jusoh (2001) choose *D* = 40 mm, 80 mm, 160 mm and *W* = 10 mm, 20 mm, 40 mm. The corresponding mean *ID* is 2.396 bits. Silfverberg et al. (2001) choose *W* = 3 mm, 6 mm, 12 mm and *D* = 25 mm, 50 mm, 100 mm, which corresponds to mean *ID* 3.269 bits. Douglas et al. (1998) used yet a different set: W= 2, 5, 10 mm and *D* = 40, 80, 160 mm, corresponding to mean *ID* 4.219 bits.

There is yet another way to calculate *TP* by averaging ID/MT ratio over all ID values in the experiment:

$$TP_c = \overline{ID / MT} = \frac{\sum_{i=1}^{N} ID_i / MT_i}{N} \qquad (7)$$

As illustrated in Figure 4, TP_c averages the inverse of the series slopes determined by ID_i / MT_i and hence is a function of the set of ID used. It therefore suffers the same problem as TP_b .

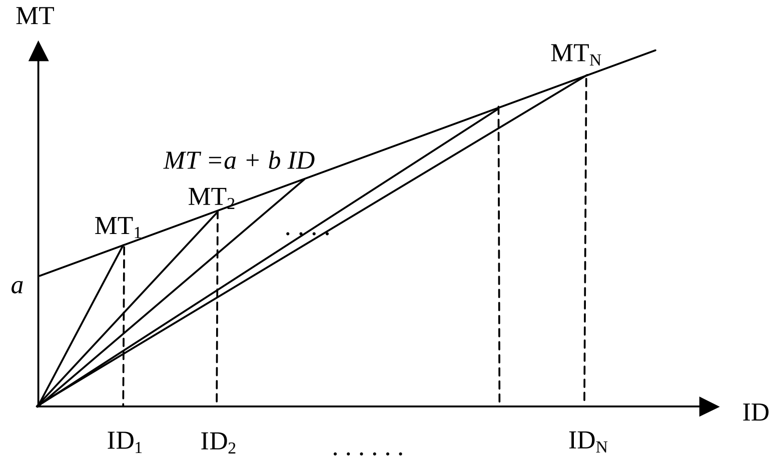

Figure 5. $TP_c = \overline{ID / MT}$ **changes as a function of ID points used.**

Sources of non-zero intercept a

The cause of the throughput problem is the non-zero Fitts' law intercept *a*. There are at least the following three possible sources of a non-zero *a.*

1. Regression error. As a statistical method, regression results cannot be perfect. Even if there is not an intrinsic constant component to *MT*, there could be a non-zero intercept due to regression noise. This is particular true when the range of *ID* used in experiments is narrow. A non-zero *a* resulted from such a cause is purely an artifact of experimental measurement. This cause can be reduced by choosing a wide enough range of *ID* in experiments.
2. A component of human visual, cognitive or motor reaction process that is independent of movement task parameters. For example the process of finding where the target is (in case of a serial, none reciprocal pointing task). The time to activate muscle movement etc.
3. Modeling error. Evidence in the literature (e.g. Crossman & Goodeve, 1963/1983, also see Fitts' original data) shows that Fitts' law may not apply to tasks with very low *ID*. When *ID* is lower than 2 bits, it is likely the task shifts towards open-loop behavior governed by Schmidt's law rather than byinformation as quantified by Fitts' index of difficulty. The movement time is hence curved in the low ID end when the horizontal scale is logarithmic. This "non-Fitts component" when forced into a linear regression on the logarithmic scale may result in a non-zero intercept.
4. A component of motor performance independent of distance or target size. For example, the time to click on a mouse button, the time to tap on the same target twice (D = 0), etc.

Category 4 is in fact a performance aspect that should not be ignored or mixed into the information (time as a function of ID) aspect of performance. It is possible that an input device is more effective than another in moving a cursor from one target to another but less so in target activation due to selection button design. When this happens, it is more informative to report it (as indicated by *a*) separately, rather than mixing it with the information aspects. In designing laptop computer pointing devices, such as touchpads or miniature in keyboard joysticks, a particular challenge is that the selection buttons cannot be well integrated into the movement action as for a mouse. Some alternative solutions have been made available in recent years, such as press-to-select or tap-to-select, which uses a vertical press or a tap on the stick or the touchpad surface to activate target selection. When researching the efficiency of these solutions, it is critical to be able to study *a* separately.

Conclusions

It is a worthy goal to find the smallest possible number of quantities that are necessary and sufficient to characterize input performance. The use of Fitts' law is a leap forward in this regard. Without Fitts' law, pointing performance cannot be generalized beyond the set of targets size and distance used. With it, input performance can be characterized by the Fitts' law parameters (*a* and *b*), plus error rate made during the experiment.

Another possible step forward is to incorporate error into Fitts' law regression by adjusting target sizes so there are a fixed percentage of trials outside of the adjusted targets. While this is a compelling method, it lacks firm theoretical or empirical foundation. We suggest both adjusted and unadjusted results with errors reported until a firm foundation is established.

A further step towards performance parameter reduction is to concentrate on a single metric – throughput. Recommended by ISO 9241-9 FDIS and adopted by an increasing number of input studies, it is unfortunately an "overshoot" on the topic of input performance characterization. The analysis presented in this paper shows that throughput is either limited to one aspect of performance, as defined by $TP_b = 1/b$, or ill-defined with value extrinsic to the input system, as calculated by $TP_a = \overline{ID} / \overline{MT}$.

The forgoing analysis points out that input performance can be a multi-dimensional property, along at least an information (*ID*) independent dimension as measured by *a*, and an information dependent dimension as measured by *b*. How the different dimensions weigh in overall performance depends on the application tasks so that trade-off decision should be made by the end-user of the performance information. If these dimensions are combined arbitrarily (e.g. incidentally by the choice of mean *ID* used in experiments), important information is lost and the result can be misleading[5].

We conclude Fitts' law in its complete form (with *a* and *b*[6]) plus error rates should be used in characterizing input performance, as these parameters are theoretically intrinsic to the input system studied and hence generalizable beyond the target parameters used in experiments.

It should also be pointed out that Fitts' law only models pointing tasks. Performance characterization using the Fitts' paradigm should be interpreted as such. Depending on its application, a study may also use other paradigms of evaluation, such as path steering or goal crossing whose movement regularities have been recently reported ((Accot & Zhai, 1997), (Accot & Zhai, 2002)). Note that although much smaller relative to the total completion time, the law of steering may also has a non-zero intercept.

---

[5] It should be pointed out that the cited studies based on ISO 9241-9 should still have internal validity between the experimental conditions within these studies. However the performance measure embodied in *TP* cannot be generalized beyond these studies as *TP* is not intrinsic. Furthermore many of them no longer report complete Fitts' law results (*a, b*, error rate, a goodness of fit), so it is not possible to recover the intrinsic *a b* measures.

[6] Or it can be said as TP_b plus *a*.

References

Accot, J., & Zhai, S. (1997). Beyond Fitts' Law: Models for Trajectory-Based HCI Tasks. In *Proc. of CHI'97: ACM Conference on Human Factors in Computing Systems*.

Accot, J., & Zhai, S. (2001). Scale effects in steering law tasks. In *Proc. of CHI 2001: ACM Conference on Human Factors in Computing Systems*.

Accot, J., & Zhai, S. (2002). More than dotting the i's - foundations for crossing-based interfaces. In *Proc. of CHI: ACM Conference on Human Factors in Computing Systems*, Minneapolis, Minnesota.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate controlled isometric joystick, step keys and text keys for text selection on a CRT. *Ergonomics, 21*, 601-613.

Crossman, E. R. F. W., & Goodeve, P. J. (1983). Feedback control of hand-movement and Fitts' law. *Quarterly Journal of Experimental Psychology, 35A*, 251-278. (First appeared in 1963).

Douglas, S., Kirkpatrick, A., & MacKenzie, S. (1998). Testing Pointing Device Performance and User Assessment with the ISO 9241, Part 9 Standard. In *Proc. of CHI'98: ACM Conference on Human Factors in Computing Systems*.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381-391.

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology, 67*, 103-112.

Guiard, Y. (2001). Distangling relative from absolute amplitute in Fitts law experiments. In *Proc. of Extended Abstracts(short talks), CHI 2001: ACM Conference on Human Factors in Computing Systems*, Seattle.

MacKenzie, I. S. (1989). A note on the information-theoretic basis for Fitts' law. *Journal of Motor Behavior, 21*, 323-330.

MacKenzie, I. S. (1992). Fitts’ law as a research and design tool in human computer interaction. *Human Computer Interaction, 7*, 91-139.

MacKenzie, I. S., & Jusoh, S. (2001). An evaluation of two input devices for remote pointing. In *Proc. of the Eighth IFIP Working Conference on Engineering for Human-Computer Interaction - EHCI 2001*, Heidelberg, Germany.

MacKenzie, I. S., Sellen, A., & Buxton, W. (1991). A comparison of input devices in elemental pointing and dragging tasks. In *Proc. of CHI’91: ACM Conference on Human Factors in Computing Systems*, New Orleans, Lousiana.

McGuffin, M., & Balakrishnan, R. (2002). Acquisition of Expanding Targets. In *Proc. of CHI: ACM Conference on Human Factors in Computing Systems*.

Meyer, E. D., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. K. (1988). Optimality in Human Motor Performance: Ideal Control of Rapid Aimed Movements. *Psychological Review, 95*(3), 340-370.

Mottet, D., Guiard, Y., Ferrand, T., & Bootsma, R. J. (2001). Two-Handed Performance of a Rhythmical Fitts Task by Individuals and Dyads. *Journal of Experimental Psychology: Human Perception and Performance, 27*(6).

Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., & Long, A. C. (2002). Interacting At a Distance: Measuring the Performance of Laser Pointers and Other Devices. In *Proc. of CHI: ACM Conference on Human Factors in Computing Systems,* Minneapolis, Minnesota.

Oh, J.-Y., & Stuerzlinger, W. (2002). Laser Pointers as Collaborative Pointing Devices. In *Proc. of Graphics Interface*.

Schmidt, R. A., Zelaznik, H., Hawkins, B., Frank, J. S., & Quinn, J. T. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review, 86*(5), 415-451.

Silfverberg, M., MacKenzie, I. S., & Kauppinen, T. (2001). An isometric joystick as a pointing device for handheld information terminals. In *Proc. of Graphics Interface,* Toronto, Canada.

Soukoreff, W., & MacKenzie, I. S. (1995). Theoretical upper and lower bounds on typing speeds using a stylus and keyboard,. *Behaviour & Information Technology, 14*, 379-379.

Welford, A. T. (1968). *Fundamentals of Skill*. London.

Zhai, S., Hunter, M., & Smith, B. A. (2000). The Metropolis Keyboard - an exploration of quantitative techniques for virtual keyboard design. In *Proc. of The 13th Annual ACM Symposium on User Interface Software and Technology (UIST)*, San Diego, California.

Zhai, S., Morimoto, C., & Ihde, S. (1999). Manual And Gaze Input Cascaded (MAGIC) Pointing. In *Proc. of CHI'99: ACM Conference on Human Factors in Computing Systems*.

Zhai, S., Smith, B. A., & Hunter, M. (2002a). Performance Optimization of Virtual Keyboards. *Human-Computer Interaction, 17*(2,3), 89-129.

Zhai, S., Sue, A., & Accot, J. (2002b). Movement Model, Hits Distribution and Learning in Virtual Keyboarding. In *Proc. of CHI: ACM Conference on Human Factors in Computing Systems,* Minneapolis, Minnesota.