Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Int. J. Human-Computer Studies 61 (2004) 791–809
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Characterizing computer input with Fitts' law parameters—the information and non-information aspects of pointing

Shumin Zhai*

IBM Almaden Research Center, 650 Harry Road, NWE-B2, San Jose, CA 95120, USA

Available online 2 November 2004

Abstract

Throughput (*TP*), also known as index of performance or bandwidth in Fitts' law tasks, has been a fundamental metric in quantifying input system performance. The operational definition of *TP* is varied in the literature. In part thanks to the common interpretations of International Standard ISO 9241-9, the "Ergonomic requirements for office work with visual display terminals—Part 9: Requirements for non-keyboard input devices", the measurements of throughput have increasingly converged onto the average ratio of index of difficulty (*ID*) and trial completion time (*MT*), i.e. $TP=ID/MT$. In lieu of the complete Fitts' law regression results that can only be represented by both slope (*b*) and intercept (*a*) (or $MT=a+b\ ID$), *TP* has been used as the sole performance characteristic of input devices, which is problematic. We show that *TP* defined as ID/MT is an ill-defined concept that may change its value with the set of *ID* values used for the same input device and cannot be generalized beyond specific experimental target distances and sizes. The greater the absolute value of *a* is, the more variable *TP* ($=ID/MT$) is. ID/MT only equals a constant $1/b$ when $a = 0$. We suggest that future studies should use the complete Fitts' law regression characterized by (a, b) parameters to characterize an input system. *a* reflects the non-informational aspect and *b* the informational aspect of input performance. For convenience, $1/b$ can be named as throughput which, unlike ID/MT, is conceptually a true constant.

Keywords: Computer input; Fitts' law; Motor control performance; Throughput; Bandwidth; Index of performance (*IP*); ISO 9241-9

*Tel.: +1 408 927 1112; fax: +1 408 927 3033.
E-mail address: zhai@almaden.ibm.com (S. Zhai).

1071-5819/$ - see front matter
doi:10.1016/j.ijhcs.2004.09.006

1. Introduction

The theoretical argument, empirical analysis, and critical literature review presented in this paper are concerned with a basic concept in Fitts' law studies, namely throughput. My main logic is mathematically obvious and can be briefly stated as the following: If a dependent variable y is a function of an independent variable x in the following way:

$$y = c_1 + c_2 x, \tag{1}$$

where c_1 and c_2 are two non-zero constants, this relationship cannot be further reduced to another constant that still captures the same regularity as modeled by Eq. (1). Both the quantity

$$u = \frac{\bar{x}}{\bar{y}} = \frac{\bar{x}}{c_1 + c_2 \bar{x}}, \tag{2}$$

and the quantity

$$v = \overline{x/y} = \frac{1}{N} \sum_{i=1}^{N} \frac{x_i}{y_i} \tag{3}$$

depend on the choice of the independent variable values $\{x_i\}$ or their average $\bar{x}$. In other words, u or v are not constants. Fitts' law studies should therefore report and rely on both c_1 (intercept) and c_2 (slope) of the Fitts' regression results.

Logically little more needs to be said in this paper. However, the misunderstanding of this simple logic is widespread (and increasingly so), often by some of the most eloquent and persuasive students of Fitts' law in the human computer interaction (HCI) literature. Furthermore, this logic pertains to an ISO standard, or the interpretations of it. These factors warrant a more thorough analysis of the throughput problem in the context of the literature, starting with the very motivation of using Fitts' law in HCI. Although the current discussion is mainly situated in HCI applications of Fitts' law, it is also relevant to Fitts' law studies in general.

2. Fitts' law studies in HCI

One important aspect of computer input research is to measure and characterize the performance of various input systems. Given the great potential diversity of input devices, such as the mouse, joystick and touchpad (or different versions of the same type of device), a critical need is to be able to compare and characterize them from a human performance perspective.

A *tempting but naïve* approach to measure input performance could be as follows: let users perform an input task such as pointing, under varied conditions such as different target size and distance; measure and record users' completion time; take the average of all trials for each input device as the performance metric; compare the values (e.g. 800 vs. 900 ms) between different devices; and make a conclusion based on these values (e.g. A is 10% faster than B on average). This seemingly natural and reasonable approach is unfortunately naïve and cannot be generalized, because such a measurement is a function of the experimental settings, in particular, the sizes and distances of the pointing targets. The conclusion based on such an approach is only valid with the set of targets tested. If the target sizes and distances were different, the same conclusion might not hold. In other words, average completion time is not only a function of the intrinsic properties of the input devices but also a function of the experimental task parameters extrinsic to the input devices.

Card, English and Burr first realized that human performance models could be used to generalize studies on input devices (Card et al., 1978). In particular, they applied the well-known Fitts' law (Fitts, 1954; Fitts and Peterson, 1964; Fitts and Radford, 1966) as a framework to study computer pointing tasks

$$MT = a + b\,ID, \tag{4}$$

where MT[1] is the mean (or expected) completion time of a pointing trial and $ID = \log_2(2D/W)$ is the index of difficulty as originally defined by Fitts (Fitts, 1954; Fitts and Peterson, 1964). D and W are target distance and size, respectively, and a and b are empirically determined constants, reflecting the performance of the input system tested.[2]

The power of Fitts' law in this context is that the infinite number and range of possible target sizes and distances can now 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 could be generalized to other target sizes and distances not used in the experiment.[3]

Since Card et al.'s work, Fitts' law has been used as a framework in the majority of input device research (many are cited elsewhere in this paper). Beyond input device comparison, Fitts' law has also been used in, for example, research of alternative interaction techniques (Kabbash and Buxton, 1995; McGuffin and Balakrishnan, 2002; Zhai et al., 2003), new interaction paradigms such as crossing (Accot and Zhai, 2002), telerobotics (Drascic, 1991), collaborative haptic devices (Sallnas and Zhai, 2003) and stylus keyboard optimization, either manually (Lewis et al., 1992; Soukoreff and MacKenzie, 1995; MacKenzie and Zhang, 1999) or algorithmically (Zhai et al., 2002a).

[1] MT originally and literally meant movement time, as opposed to reaction time (RT) (Fitts and Peterson, 1964), particularly in "discrete" pointing tasks. In the literature of HCI, however, the separation of MT and RT is more difficult. Consistent with the majority of Fitts' law studies in HCI, we use MT as the entire "trial completion time", unless explicitly stated otherwise.

[2] An obvious problem with the definition of $ID = \log_2(2D/W)$ is that when D, the distance from the center of one target to another, is zero, ID (and hence T) tends to negative infinity. 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. $ID = \log_2((D+W)/W)$, which resolves the negative ID problem. See also Welford's formulation (Welford, 1968).

[3] This is only true within a reasonable range, but not to extremely large- or small-scale ends.

3. The use of throughput or index of performance

While there is no doubt Fitts' law as a general framework has been tremendously successful, many of the details in the actual application of it are a matter of ad hoc choice, debate and discussion (Fitts and Peterson, 1964; MacKenzie, 1992; Isokoski and Raisamo, 2002). A critical one is the choice of Fitts' law parameters as performance metrics. In particular, implicitly or explicitly, it has been desired to use one parameter, namely the index of performance, more commonly referred to as throughput (*TP*) in recent HCI literature, to quantify pointing performance. How I_p is defined is a source of confusion and a critical problem, as this paper shows. Fitts himself was not completely consistent in this regard. He initially defined

$$I_p = ID/MT, \tag{5}$$

as the metric to quantify motor performance (adapted from Eq. (2) of Fitts, 1954). Note that in Fitts' original definition I_p was dependent on a specific *ID* point. More precisely, I_p was defined on a specific point (the *i*th sample) on the regression line

$$I_p(i) = ID(i)/MT(i). \tag{6}$$

Fitts and Peterson used the same quantity (*ID* to *MT* ratio) as a measure ‘‘analogous to man's capacity for executing a particular class of motor responses in bits per second’’ (Fitts and Peterson, 1964). Later, Fitts and Radford avoided the reference to this definition of I_p (Fitts and Radford, 1966). Instead, they referred to the *slope constant* as the ‘‘relatively constant information capacity over a range of movement conditions’’ (Fitts and Radford, 1966, p. 476). To put it more explicitly, they effectively defined I_p as

$$I_p = 1/b. \tag{7}$$

By this definition I_p is independent of the measurement point on the regression line, since it is the slope of a straight line. This definition of index of performance is only a matter of naming convenience, for it (inversely) corresponds to *b*, with no substantive difference from the concept of *b*.

In the same paper (p. 480), Fitts and Radford also referred to both *a* and *b* parameters for assessing human motor control performance: ‘‘The rate of information processing by the human motor system, as inferred from the intercept and slope constants ..., appears to remain relatively constant ...’’.

Fitts' central thesis was that the human motor capacity was relatively constant across different experimental conditions. To prove such a proposition, the difference between the two definitions of I_p was not critical since the difference in I_p compared to the wide range of *ID* change was small enough in either definition. For example, although I_p as defined by Eq. (6) changed from 8.92 to 12.57 bits/s in Fitts (1954), well over 20%, Fitts could still conclude that the information processing capacity in the motor system is ‘‘relatively’’ constant, in comparison to the many folds of distance and width change.

In HCI research, particularly when Fitts' law is used as a tool for device performance comparison, such a difference becomes much more crucial. In practice, HCI researchers have differed in their choice of I_p definition. For example, some studies (Card et al., 1978; MacKenzie, 1991; MacKenzie et al., 1991; Soukoreff and MacKenzie, 1995; Zhai et al., 1999 used Eq. (7). Most recent studies Douglas et al., 1998; MacKenzie and Jusoh, 2001; Silfverberg et al., 2001; Myers et al., 2002; Oh and Stuerzlinger, 2002) use Eq. (5).

ISO 9241-9 (ISO, 2000), in particular its "informative" Annex B (as opposed to "normative" sections), attempts to provide a standard approach for input device evaluation based on Fitts' law. It recommends using throughput as the sole metric of input device performance metric

$$TP = ID_e/MT, \tag{8}$$

where ID_e is the effective index of difficult adjusted by using effective width of the target W_e which is defined as 4.133σ, where σ is the standard deviation of the actual endpoint distribution.

The interpretation of Eq. (8) can be ambiguous. Implicitly, ISO 9241-9 assumes zero intercept of Fitts' law regression ($a = 0$) (Fig. 1). When the intercept a is not zero (which is the general case, to be discussed later), the value of TP depends on the x-coordinate on the Fitts' law regression line (Fig. 4). Different TP values will be obtained at different points of ID.

One interpretation of ISO 9241-9 is to use the ratio of means: "According to ISO 9241-9, throughput is obtained from the division of means (see Eq. (1)), not from the slope reciprocal in a regression model" (MacKenzie and Jusoh, 2001, p. 245). More explicitly, in the rest of this paper we label this definition throughput as TP_a:

$$TP_a = \overline{ID}/\overline{MT}. \tag{9}$$

In contrast, ISO 9241-9 Annex B in fact also states, "The slope of the line represents the throughput of the device, in bits per second" (ISO, 2000). In other words, it implies another definition that we hereafter designated as TP_b:

$$TP_a = 1/b. \tag{10}$$

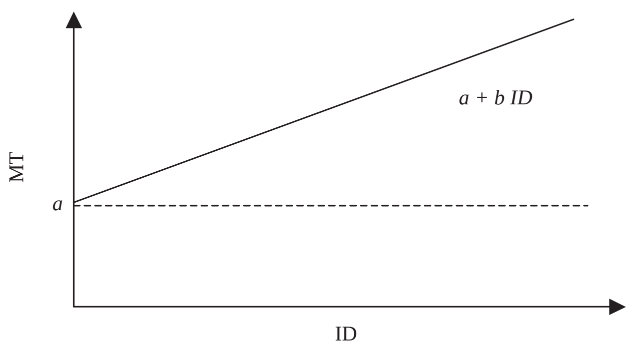

Fig. 1. Fitts' law: movement time in pointing increases linearly with ID, in addition to the constant a.

Another possible interpretation of Eq. (8) is to calculate *TP* by averaging *ID* values measured at different points of *ID* (e.g. MacKenzie and Soukoreff, 2003)

$$TP_c = \overline{TP_i} = \frac{1}{N}\sum_{i=1}^{N} ID_i/MT_i. \tag{11}$$

The goal of an ISO definition of throughput is undoubtedly desirable. First, it standardizes input device evaluation. Second, it uses one parameter to capture the entire quality of an input device. Unfortunately, TP_a or TP_c, the two more common interpretations of ISO (Douglas et al., 1998; MacKenzie and Jusoh, 2001; Silfverberg et al., 2001; Myers et al., 2002; Oh and Stuerzlinger, 2002), are ill-defined metrics in the sense that they change with choices of *ID* values used in an experiment.

Now let us examine more specifically the problem of TP_a. $TP_a = \overline{ID}/\overline{MT}$ incorporates both *a* and *b* parameters of Fitts' law (Fig. 1) into one metric. By merging the non-zero *a* with *b* TP_a is no longer a value intrinsic to an input system. Since *MT* follows Fitts' law, we may substitute *MT* with $a+b\ ID$ in Eq. (9) and obtain

$$TP_a = \overline{ID}/(a + b\overline{ID}). \tag{12}$$

Eq. (12) shows that TP_a depends not only on *a* and *b,* but also on the mean *ID* used in the experiment. Fig. 2 illustrates how TP_a changes with $\overline{ID}$ at typical Fitts' law parameter values $b = 0.125$ (s/bit) and $a = 0.1$, 0.2 or 0.4 s. Depending on $\overline{ID}$, TP_a moves up and down on the curves in Fig. 2.

The sensitivity of TP_a to $\overline{ID}$ varies with *a.* The greater *a* is, the more slowly TP_a converges to $1/b$. When $a = 0$, $TP_a = 1/\mathrm{b}$.

TP_a has an asymptote of $1/b$ when $\overline{ID}$ increases toward infinity. This means that a more stable value of TP_a can be measured if a greater $\overline{ID}$ is used in the experiment.

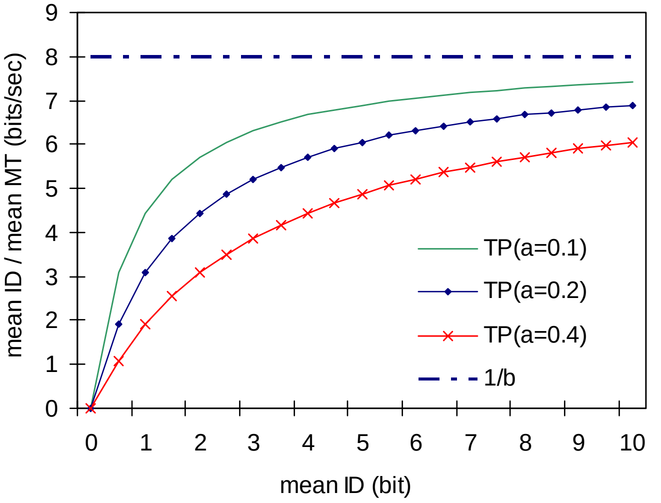

Fig. 2. Throughput defined by $TP_a = \overline{ID}/\overline{MT}$ is a function of mean ID.

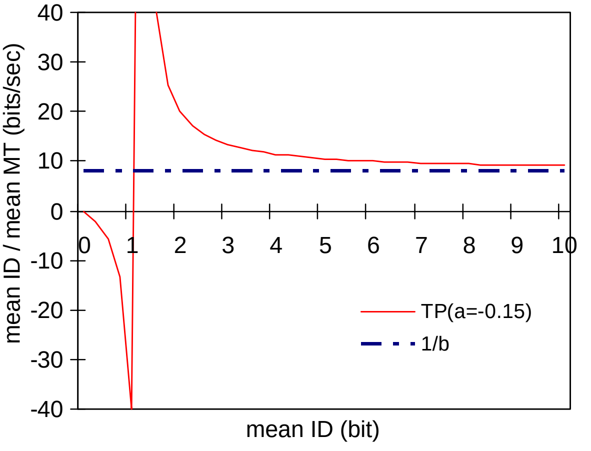

Fig. 3. $TP_a = \overline{ID}/\overline{MT}$ as a function of mean ID, when a is negative.

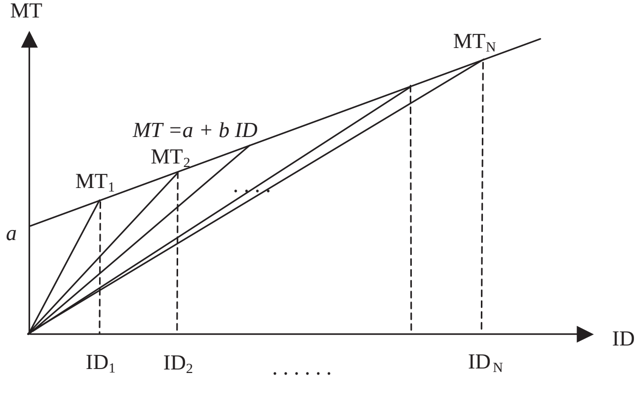

Fig. 4. $TP_c = \overline{ID/MT}$ changes as a function of ID points used.

On the other hand, this is totally unnecessary because the asymptote can be easily estimated from Fitts' law regression ($1/b$).

Often a negative a value can be found in an experiment (see numerous instances in, e.g. Fitts and Radford, 1966; MacKenzie, 1991). When a is negative, TP_a estimation is even more unstable. Not only TP_a is a variable, it also has a singularity point at $\overline{ID} = |a|$, where TP_a is infinity. Fig. 3 shows TP_a as a function of mean ID when $a = -0.15$ s.

TP_c suffers a similar problem as TP_a. As illustrated in Fig. 4, TP_c averages all ID_i/MT_i and hence is a function of the set of ID_i used. As we can see, in the presence of a positive a, the higher the ID point is, the lower the MT/ID ratio is. The average of these ratios depends on the range of ID values chosen. For exactly the same input device, if a set of trials were collected at an ID value higher than ID_N (Fig. 4), a lower average TP_c would be obtained.

In summary, for the same device both TP_a and TP_c may change with the experimental setup (target sizes and distances). This diminishes the hope that "As a result (of conforming to ISO 9241-9), a payoff is now appearing. The payoff is the ability to compare results across studies with confidence that the comparison is 'apples with apples'" (MacKenzie and Jusoh, 2001, pp. 245–246).

4. An empirical example

To illustrate the difference and confusion that can be caused by different measurements of *TP*, we select data from Card et al. (1978), the first Fitts' law application in input device evaluation, as an empirical test case.

Card, English and Burr experimented with four devices under five target distances (1, 2, 4, 8, 16 cm) and four target widths (1, 2, 4, 10 character). Each character was 0.246 cm wide and 0.456 cm high. The 20 target conditions formed 12 unique *ID* values. Card and colleagues summarized their results in Fitts' law regression plots and corresponding equations. For the mouse, which was still in its formative days, the Fitts' law regression result was

$$MT = 1.03 + 0.096\log_2(D/W + 0.5). \tag{13}$$

Card and colleagues used the TP_b definition of throughput, which is the best definition as this paper shows, and estimated its value at

$$TP = 1/0.096 = 10.3\,\text{bits/s}. \tag{14}$$

Recently, MacKenzie and Soukoreff (2003, hereafter M&S) updated Card's and colleagues' results according to many of the recommendations of ISO and "the current practice" in the following aspects: the use of min(W, H) as width for 2D targets,[4] the use of effective width rather than nominal width, treatment of error (which could not be revised due to unavailability of data), the Shannon rather than Welford definition of *ID*, and using the TP_c rather than TP_b definition of throughput. Using the TP_c definition, M&S re-estimated the throughput of the mouse used in Card's and colleagues' experiment as 2.65 bits/s (the mean of Column 8 in Table 1). As M&S stated, this is substantially different from Card's and colleagues' estimation.

Card *et al.*'s data and M&S's recalculation can dramatically demonstrate the dependency of TP_c on the set of *ID* values under which experimental data were collected. Suppose Card and colleagues did not collect data on the condition of

[4]This author could not find min(W, H) recommendation of 2D target size in ISO. Instead it notes that "For a selection, pointing or dragging task, the target width is measured along the direction of movement." (p. 29). (MacKenzie and Buxton, 1992) gives a systematic early account on the use of *min(W, H)*. Without affecting the essence of the current *TP* analysis and to be consistent with M&S, we use *min(W, H)* to calculate *ID* in the current analysis, although this method of calculating Card's and colleagues' data yielded poor goodness of fit ($R^2 < 0.8$). How 2D target size should be determined requires further analysis, but its result will not change the essence of the current discussion. See (Accot and Zhai, 2003) for a recent systematic treatment and our view of Fitts' law formulation when a target has two dimensional constraint.

Table 1
Card's and colleagues' data in 1-oz stylus-tapping experiment (adapted from MacKenzie and Soukoreff, 2003)

1 D (cm)	2 W (char)	3 $W1{:}1D$ width (cm)	4 $ID=(D/W1+0.5)$ "Welford"	5 $W2=\min(W,H)$	6 $ID=(D/W2+1)$ "Shannon"	7 MT (s)	9 ID/MT
1	1	0.246	2.19	0.246	2.34	1.24	1.88
1	2	0.492	1.34	0.456	1.67	1.18	1.42
1	4	0.984	0.6	0.456	1.67	1.11	1.5
1	10	2.46	−0.14	0.456	1.67	0.97	1.73
2	1	0.246	3.11	0.246	3.19	1.34	2.38
2	2	0.492	2.19	0.456	2.43	1.24	1.96
2	4	0.984	1.34	0.456	2.43	1.18	2.06
2	10	2.46	0.39	0.456	2.43	1.06	2.28
4	1	0.246	4.07	0.246	4.11	1.45	2.83
4	2	0.492	3.11	0.456	3.29	1.34	2.46
4	4	0.984	2.19	0.456	3.29	1.24	2.65
4	10	2.46	1.09	0.456	3.29	1.13	2.91
8	1	0.246	5.05	0.246	5.07	1.53	3.31
8	2	0.492	4.07	0.456	4.21	1.45	2.9
8	4	0.984	3.11	0.456	4.21	1.34	3.15
8	10	2.46	1.91	0.456	4.21	1.27	3.31
16	1	0.246	6.03	0.246	6.05	1.66	3.64
16	2	0.492	5.05	0.456	5.17	1.53	3.38
16	4	0.984	4.07	0.456	5.17	1.45	3.56
16	10	2.46	2.81	0.456	5.17	1.37	3.77

Column 7 is M&S's estimates from Card et al.'s scatter plot. Columns 5 and 6 were M&S's recommendation.

$D = 1$ cm (removing the top four rows in Table 1), TP_c would have been 2.91 bits/s. If Card and colleagues did not collect data on $D = 16$ cm (removing the bottom four rows in Table 1), TP_c would have been 2.42 bits/s. Shifting distance condition by one level in the experiment, a 20% difference of TP_c would have been reported, while everything else, in the particular the device itself, remained the same.

Similarly, using the TP_a definition of throughput will also yield variable values depending on the experimental conditions included. Based on the means of Column 7 (*MT*) and Column 6 (*ID*) in Table 1, TP_a is 2.72, 2.95 or 2.47 bits/s, respectively, when all *D* conditions are included, when $D = 1$ cm is excluded or when $D = 16$ cm is excluded in the calculation.

In contrast, if we use the *entire* Fitts' law representation (*a* and *b* or the regression line), we find little difference when dropping either $D = 1$ or *16* cm conditions. If we use the *ID* formulation recommended by M&S (i.e. $ID = \log 2(D/\min(W,H)+1)$ and perform linear regression between Column 7 (*MT*) and Column 6 (*ID*) of data in Table 1, we obtain the following Fitts' law results:

with all *D* conditions,

$$MT = 0.894 + 0.115ID, \tag{15}$$

without $D = 1$ cm condition,

$$MT = 0.863 + 0.122ID, \tag{16}$$

without $D = 16$ cm condition,

$$MT = 0.888 + 0.119\,ID. \tag{17}$$

Graphically (Fig. 5), we can see that dropping a set of distance condition on either end (16 or 1 cm) had little impact on the *total* picture of Fitts' law regression. It demonstrates the remarkable power of using Fitts' law (in its entirety) as a framework of input device evaluation: *a* and *b* (or the regression line) represents the intrinsic property of the input device, relatively invariant to the experimental setup.

Pointing out the substantial difference between their estimate and Card's and colleagues' original estimate of *TP*, M&S noted that "An additional issue is the high intercept in their regression line". Card et al. attributed the intercept *a* value to adjusting the hand grasp on the mouse at the beginning of a trial and the time for the target selection by button presses. "Arguably, this time should be excluded. Doing so by reducing each MT in Table 1 by, nominally, 0.5 s, increases throughput to 4.32 bps, a value typical in ISO-conforming studies for the mouse", reasoned M&S. This removal of a portion of *a* is indeed arguable, but it does not change the essence that TP_c is ill-defined. Repeating the above numeric estimates of TP_c with half of the *a* value would reveal the same unstable effect, albeit to a lesser extent.

Since the publication of the final draft of ISO 9241-9, numerous input devices studies have been conducted, using TP_a or TP_c, rather than the entire Fitts' law equations as the measure of performance. The devices studied include isometric joystick and touchpad (Douglas et al., 1998), remote control mouse (MacKenzie and Jusoh, 2001; Silfverberg et al., 2001), mouse size and shape (Isokoski and Raisamo, 2002) and laser pointers (Myers et al., 2002; Oh and Stuerzlinger, 2002). Using TP_a

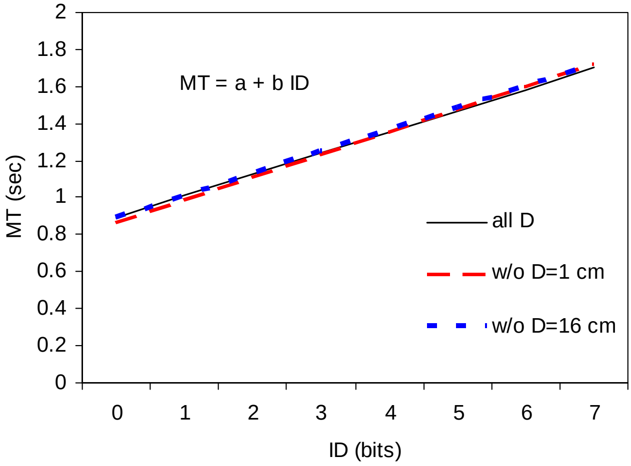

Fig. 5. Fitts' law regression results of under different distance conditions based on Card et al. (1978) and MacKenzie and Soukoreff (2003). There was little difference between the conditions if both parameters of Fitts' law are used as illustrated graphically here, in contrast to the large differences of TP_a or TP_c (see text).

or TP_c has been considered the "current practice" by some researchers (e.g. MacKenzie and Soukoreff, 2003). It would be informative to do the same kind of analysis as to Card et al.'s study of these studies. With the exception of Isokoski and Raisamo (2002), who reported Fitts' law regression plots (but not the equations), most of these studies only reported TP values (as recommended or interpreted from ISO 9241-9). As a result, we cannot know what the entire pictures of Fitts' law regression are in these studies, since the same TP value (by the definition of TP_a or TP_c) can be resulted from an infinite set of a and b values, due to averaging effect. If a is or is close to being zero, the TP values reported in these studies will be close to $1/b$ and hence can be generalized. Otherwise, their values are tied with the set of target parameters used in these studies.

5. Non-zero intercept *a*

Clearly, the problem of throughput has to do with the non-zero Fitts' law intercept a. If the Fitts' law regression intercept has a zero or near-zero intercept ($a = 0$), all three definitions of throughput would converge and will all be conceptually and quantitatively equivalent to b. Some researchers and advocates of throughput in the form of TP_a or TP_c have certainly realized this. "All else being equal, the two calculations should yield reasonably similar results provided the intercept is zero, or close to zero" (MacKenzie and Jusoh, 2001, p. 245). When $a = 0$ the only parameter left in the Fitts' law equation is b. Defining TP as $1/b$ would only be a matter of naming convenience. There would not be a need to define TP_a or TP_c.

When a is not zero, TP_a or TP_c are ill-defined, as we have shown. It is therefore important to understand the sources of a non-zero intercept. There are two broad and non-exclusive categories of explanations for a non-zero intercept in Fitts' law. One has to do with the fact that MT may include components that are insensitive to information (as quantified by ID) and the other has to do with errors in the slope estimation. More specifically, there are at least the following *possible* sources of a non-zero a.

(1) *Regression error*: As a statistical method, regression results are subject to noise in data. Even if there is not an intrinsic constant component to MT, there could be a non-zero intercept due to regression noise. Obviously, the narrower the range of ID used in an experiment, the less robust the MT vs. ID regression, and hence the less reliable the estimate of a. A non-zero a resulted from such a cause is purely an artifact of experimental measurement. This cause may be reduced by careful filtering of noise, such as outliers caused by instrumental error and choosing a wide-enough range of ID in experiments.

(2) *The use of nominal* vs. *effective target width*: Another source that may contribute to a non-zero a is related to the measurement of target width. The idea of using the actual spread of the hits dispersion (4.133σ, where σ is the standard deviation of the hits distribution along the movement direction), rather than specified (nominal) target width, emerged soon after Fitts' original publication (e.g. Crossman, 1960). It can even be traced back long before Fitts' law (Woodworth, 1899), as pointed out by Fitts and Radford (1966). Although the idea of effective width has been around for such a long period, its theoretical justification (if a post hoc measurement is cognitively equivalent to a priori task specification) and empirical validity (if human performance adjusted with effective width is equivalent to a nominal width of the same magnitude, had the performer followed the exact specification) still calls for a thorough investigation (Zhai, Kong and Ren, 2004). Nonetheless, clearly the use of nominal vs. effective width may change the value of a. For example, Fitts and Peterson (1964) made calculations of both W and W_e and found that a changed from -63 to $+8$ ms. MacKenzie (1991, Table 2, Chapter 2) recalculated the data from Fitts (1954) and found the value of a changed from 12.8 to -73 ms in the case of 1 oz stylus and -6.2 to -118 ms in the case 1 lb stylus. A relevant phenomenon can be inferred from Guiard's analysis of Fitts' index of difficulty (Guiard and Ferrand, 2002): when W is large (low ID, to the left side of a Fitts' plot), performers tend to under-use the given size (slower than they should be, pushing the left side of Fitts law regression line up). When W is small (high ID, to the right side of a Fitts' plot), performers tend to over-use the given size (faster than they should be, pushing the right side of Fitts' law regression line down). This means the Fitts' law line is higher on the left side and lower on the right side than it "should be", causing a greater intercept a. See Zhai et al. (2004) for a recent and systematic treatment of this topic.

(3) *Modeling error*: Evidence in the literature (Crossman and Goodeve, 1983 (Original Report 1963)) 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 that can be better described by Schmidt's law, a linear model between spatial variance and time (Schmidt et al., 1979), rather than by information 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' law component", when forced into a linear regression on the logarithmic scale, may result in a non-zero intercept.

(4) *A component of human visual, cognitive or motor reaction/activation process that is independent of movement task parameters*: For example, the process of finding where the target is, which can be explicitly and separately modeled as reaction time in case of a serial, non-reciprocal pointing task (Fitts and Peterson, 1964), is certainly not zero. Although slightly increasing with *ID*, reaction time tends to be rather constant (Fitts and Peterson, 1964; Zhai et al., 2003). When such a reaction time is not explicitly measured in HCI applications, they are lumped into a non-zero a.

(5) *A component of motor performance independent of distance or target size*: Obviously, in a computer input system, the time to click on a mouse button, the time to tap on the same key in stylus keyboard twice (MacKenzie and Zhang, 1999; Zhai et al., 2002b), is not likely to be related to Fitts' index of difficulty and should be captured by the value of a. The same is true of the "hand-grasping adjustment" time in Card et al.'s text-editing task (Card et al., 1978). Conceivably, some form of Fitts' law tasks may have a greater time component that is independent of distance and target size. For example, in the disc transfer task used in Fitts (1954), it should take certain amount of time to take a disc off the shaft, regardless of the distance to be traveled to another shaft. Indeed, the a values in Fitts' disc transfer task were estimated at 150 or 223 ms depending on *ID* definition (see Table 2, Chapter 2 in MacKenzie, 1991), which were much greater than the a values (12 or -37 ms) of the stylus-tapping task in the same study.

In summary, the Fitts' law regression intercept a can be non-zero for various reasons. Some, such as category 1, are due to measurement and modeling noise and should be minimized as much as possible in order to obtain an accurate estimate of a (and b). Others, such as category 5, are in fact a performance aspect that should be explicitly measured and stated.

a can also be very close to 0, particularly in reciprocal stylus-tapping tasks (two targets tapped repeatedly back and forth). In reciprocal stylus tapping, little cognitive or visual reaction time is involved, except perhaps in the first trial in a block. There is no button click involved for the purpose of selection at the end of a trial either. It is quite likely a is or is close to zero in such tasks. See data in Fitts (1954) and Fitts and Peterson (1964), its reanalysis in MacKenzie (1991) and many of the experiments in Zhai et al. (2004).

In any case, the use of TP_a and TP_c is either unnecessary (when $a = 0$, b is the only parameter left) or misinforming (when a is non-zero).

6. Remedies and the information and non-information aspects of pointing

The foregoing analysis suggests that the definition of TP_a or TP_c and the use of throughput in input device comparison are problematic. There are three possible remedies to these problems: (1) standardization of *ID* set in testing; (2) return to two-parameter Fitts' law modeling; and (3) a clear exclusion of any non-information aspect of pointing from Fitts' law modeling.

Given that the value of TP_a and TP_c changes with *ID* set. The first possible solution to the problem is to standardize *ID* values in pointing device research to, for example, 2, 4, 6, 8 bits, which would cover most of the pointing movements actually made in current GUIs. Once the *ID* set is standardized, device measurements based on TP_a, TP_b or TP_c can all be compared across studies, as long as their definitions are clearly stated. There are some practical difficulties to this approach, because the range of *ID* in an experiment often depends on the practical application that motivates the study. For example, the range of movement in stylus keyboarding is narrower than those in menu selection on a GUI interface, given that the soft keys in a stylus keyboard are tightly packed. Zhai et al. (2002b) had to conduct specific Fitts' law estimations due to this consideration. In practice, same researchers could change the set of *ID* values in different studies. For example, MacKenzie and Jusoh (2001) choose $D = 40, 80, 160$ mm and $W = 10, 20, 40$ mm. The corresponding mean *ID* is 2.40 bits. Silfverberg et al. (2001) choose $W=3$, 6, 12 mm and $D=25$, 50, 100 mm, which corresponds to mean *ID* 3.27 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* 4.22 bits. More fundamentally, the very motivation of using Fitts' law in input device evaluation is to be able to generalize evaluation results beyond a specific set of experimental settings. If we used a standard and a fixed set of *ID* value, we may as well use the *tempting but naïve* approach of reporting the more intuitive average task completion time under a standard set of targets, as described earlier in the paper.

The second remedy is to return to the complete formulation of Fitts' law by its two parameters (*a, b*) representation. *b*, being the coefficient of *ID* (in bits) to time (s or ms), can be viewed as the informational aspect of pointing performance (in the unit of s/bit). It indicates the temporal investment an input system takes to express each additional bit of information. Merely for the sake of convenience, TP_b the inverse of *b*, in unit of bits/s, can be referred to as "throughput" or index of (informational) performance. *a*, in contrast, is a constant that is independent of *ID*. It indicates the information-independent aspect of pointing performance. As illustrated by the hypothetical Fig. 6, if an input device (A) is more effective than another (B) in moving a cursor from one target to another (measured by *b* or TP_b), but less efficient in target activation due to the button or other selection aspects of its design (measured by *a*), the two Fitts' law regression lines would cross each other, indicating that the relative performance of the two devices would depend on *ID*. In that case, it is very important to report both *a* and *b*, so an informed decision can be made on the basis of the actual set of *ID* values in an actual application (weighted by their frequency of occurrence).

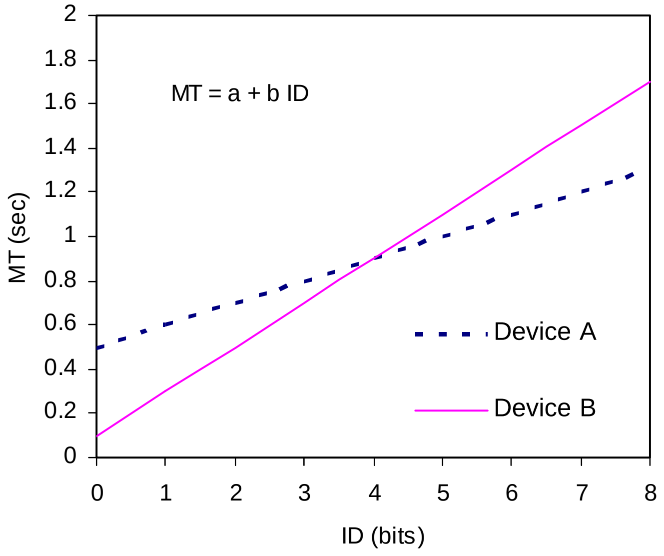

Fig. 6. A trade-off occurs when two devices have different *a* (information-independent) and *b* (information-dependent) values, depending on the *ID* value.

In designing computer pointing devices often it is important to measure and understand the non-information aspect of pointing separately. For example, the relative location, the form factor and operating finger of the activation buttons in a touchpad or miniature keyboard joystick (e.g. IBM Trackpoint$^{\mathrm{TM}}$) in laptop computers are quite different from those of the mouse. The impact of these button design factors on pointing time, which is obviously unrelated to the information aspect of pointing as quantified by *ID* therefore should be measured by the value of *a*, is interesting and informative to designers. Some alternative solutions have been attempted in recent years on these devices, such as press-to-select and tap-to-select, which use a vertical press or a tap on the stick or the touchpad surface to activate target selection. When researching these alternatives (e.g. Silfverberg et al., 2001), measuring the difference in *a* separately would give us more direct and more sensitive comparison of their impact.

The third remedy is to explicitly exclude all non-information aspects of pointing, such as visual and cognitive reaction time, memory retrieval time, motor activation time, button click time and device acquisition time, from Fitts' law modeling and treat these factors separately. If we can be certain that no *ID*- (i.e. information) independent component is left in the measurement of *MT*, any non-zero intercept in Fitts' law regression would be the result of regression or modeling noise. In that case it will be more logical to force the Fitts' law regression through the origin (0, 0) of the (*MT, ID*) coordinate. Many of today's data processing software, such as Excel$^{\mathrm{TM}}$, in fact offer this option. As a result of this approach only *b* is left in the Fitts' model, and all three throughput definitions would be equivalent and independent of *ID*; their difference will only amount to different arithmetic methods of the same quantity ($1/b$). This third approach in fact is not fundamentally different from the second approach, although it is even more explicit in treating non-information aspect of pointing separately. The advantage of this approach is that we may suppress various noises in regression instead of including them in the "catch all" a value. The danger of this approach, however, is that we may force Fitts' law into a one-parameter model, while a subtle, conceptually unobvious non-information component still exists in MT measurements.

7. Conclusions and final remark

Most students of Fitts' law, this author included, have been unaware of the pitfalls with the concept of throughput. For example, my colleagues and I first used TP as the sole parameter of Fitts' law in stylus keyboard modeling, except when the user taps on the same key twice (MacKenzie and Zhang, 1999; Zhai et al., 2002a). Later we found that a non-zero a could not be ignored in actual stylus tapping from one key to another (Zhai et al., 2002b), possibly due to a non-information component (e.g. retrieving and confirming the location of the next key in human memory). Although this did not change the relative efficiency order of various stylus keyboard layouts, the existence of distance-independent time component a did mean the efficiency percentage increase due to optimization was somewhat smaller than previously estimated (Zhai et al., 2002b).

It is a desirable 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 two Fitts' law parameters (a and b). Researchers have attempted to use throughput as a single quantity to characterize input systems, but the operational definition of throughput has not been consistent. ISO 9241-9 aimed at providing a standardized definition of TP, but the aftermath has only been more problematic to date. The value of throughput defined as the ratio of mean completion time and mean index of difficulty (TP_a) or as the mean values of ID/MT from different ID points (TP_c) depends on the specific target settings, hence defeating the very purpose of using Fitts' law in input device evaluation. If we need to use the concept of throughput or index of performance, it should be defined as a simple inversion of the information coefficient b, and be used together with the a parameter, the information-independent aspect of pointing.

We conclude that Fitts' law in its complete form (with a and b) should be used in characterizing input performance, which has been done until recently (e.g. Card et al., 1978), since these parameters are theoretically intrinsic to the input system studied and hence can be generalized beyond the target parameters used in experiments. Alternatively, we may force Fitts' law into a one-parameter (b) model by regression through the origin, if and only if we can confidently devise methods that exclude *all* information independent components in the measurement of movement time MT.

Acknowledgments

This paper was seeded in a number of casual but intense discussions and debates with Yves Guiard, Michel Beaudouin-Lafon, Pernilla Qvarfordt and other colleagues in the cafés, Metro trains, sidewalks and homes in Paris in the summer of 2002 when I was on sabbatical at the University of Paris-Sud. I thank them for providing a constructive sounding board without which this work would never have been done. I also thank my colleagues at the IBM Almaden Research Center, in particular Johnny Accot, who had realized the issues discussed in this paper long before they finally caught my attention, and Barton Smith, who has provided much input and feedback. I would also like to thank Scott Mackenzie for his many years of persevering work on Fitts' law which inspired and set the stage of this study. Finally, I gratefully acknowledge the anonymous reviewers who provided many insightful comments and suggestions that helped to improve the final version of this paper.

References

Accot, J., Zhai, S., 2002. More than dotting the *i*'s—foundations for crossing-based interfaces. Proceedings of the CHI 2002: ACM Conference on Human Factors in Computing Systems, CHI Letters 4(1). ACM, Minneapolis, MN, pp. 73–80.

Accot, J., Zhai, S., 2003. Refining Fitts' law models for bivariate pointing. Proceedings of the CHI 2003, ACM Conference on Human Factors in Computing Systems, CHI Letters 5(1). ACM, Fort Lauderdale, FL, pp. 193–200.

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., 1960. The information capacity of the human motor system in pursuit tracking. Quarterly Journal of Experimental Psychology 12, 1–16.

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

Douglas, S., Kirkpatrick, A., MacKenzie, S., 1998. Testing pointing device performance and user assessment with the ISO 9241, part 9 standard. Proceedings of the CHI'98: ACM Conference on Human Factors in Computing Systems, pp. 336–343.

Drascic, D., 1991. Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing. Proceedings of the Human Factors Society 35th Annual Meeting, San Francisco, CA, pp. 1367–1371.

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.

Fitts, P.M., Radford, B.K., 1966. Information capacity of discrete motor responses under different cognitive sets. Journal of Experimental Psychology 71, 475–482.

Guiard, Y., Ferrand, T., 2002. Does the index of difficulty measure difficulty? A note on Fitts' aimed-movement paradigm. A short version appears as Guiard, Y., Ferrand, T. (1998). Effets de gamme et optimum de difficulté spatiale dans une tâche de pointage de Fitts. Science et Motricité 34, 19–25, unpublished manuscript.

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

Isokoski, P., Raisamo, R., 2002. Speed-accuracy measures in a population of six mice. Proceedings of the APCHI 2002: 5th Asia Pacific Conference on Computer Human Interaction). Science Press, Beijing, China, pp. 765–777.

Kabbash, P., Buxton, W., 1995. The "Prince" technique: Fitts' law and selection using area cursor. Proceedings of the CHI'95: ACM Conference on Human Factors in Computing Systems. ACM, Dever, CO, pp. 273–279.

Lewis, J.R., Kennedy, P.J., LaLomia, M.J., 1992. Improved typing-key layouts for single-finger or stylus input (Technical Report No. TR 54.692): IBM Technical Report TR 54.692.

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

MacKenzie, I.S., 1991. Fitts' law as a performance model in human–computer interaction. Unpublished doctoral dissertation, University of Toronto, Toronto, Canada.

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., Buxton, W., 1992. Extending Fitts' law to two-dimensional tasks. Proceedings of the CHI'92: ACM Conference on Human Factors in Computing Systems, pp. 219–226.

MacKenzie, I.S., Jusoh, S., 2001. An evaluation of two input devices for remote pointing. Proceedings of the Eighth IFIP Working Conference on Engineering for Human–Computer Interaction—EHCI 2001. Springer, Heidelberg, Germany, pp. 235–249.

MacKenzie, I. S., Sellen, A., Buxton, W., 1991. A comparison of input devices in elemental pointing and dragging tasks. Proceedings of the CHI'91: ACM Conference on Human Factors in Computing Systems. New Orleans, LA, pp. 161–166.

MacKenzie, I.S., Zhang, S.X., 1999. The design and evaluation of a high-performance soft keyboard. Proceedings of the CHI'99: ACM Conference on Human Factors in Computing Systems, pp. 25–31.

MacKenzie, I.S., Soukoreff, R.W., 2003. Card, English, and Burr (1978)—25 years later. Proceedings of the Extended Abstracts of the ACM Conference on Human Factors in Computing Systems—CHI 2003, pp. 760–761.

McGuffin, M., Balakrishnan, R., 2002. Acquisition of expanding targets. Proceedings of the CHI 2002: ACM Conference on Human Factors in Computing Systems, CHI Letters 4(1). ACM, pp. 57–64.

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. Proceedings of the CHI 2002: ACM Conference on Human Factors in Computing Systems, CHI Letters 4(1). ACM, Minneapolis, MN, pp. 33–40.

Oh, J.-Y., Stuerzlinger, W., 2002. Laser pointers as collaborative pointing devices. Proceedings of the Graphics Interface. A.K. Peters and CHCCS, pp. 141–149.

Sallnas, E.-L., Zhai, S., 2003. Collaboration meets Fitts' law: passing virtual objects with and without haptic force feedback. Proceedings of the INTERACT 2003: IFIP Conference on Human–Computer Interaction, pp. 97–104.

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, 415–451.

Silfverberg, M., MacKenzie, I.S., Kauppinen, T., 2001. An isometric joystick as a pointing device for handheld information terminals. Proceedings of the Graphics Interface. Canadian Information Processing Society, Toronto, Canada, pp. 119–126.

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.

Woodworth, R.S., 1899. The accuracy of voluntary movement. The Psychological Review, Series of Monograph Supplements 3, 1–114.

Zhai, S., Morimoto, C., Ihde, S., 1999. Manual and gaze input cascaded (MAGIC) pointing. Proceedings of the CHI'99: ACM Conference on Human Factors in Computing Systems. ACM Press, pp. 246–253.

Zhai, S., Hunter, M., Smith, B.A., 2002a. Performance optimization of virtual keyboards. Human–Computer Interaction 17, 89–129.

Zhai, S., Sue, A., Accot, J., 2002b. Movement model, hits distribution and learning in virtual keyboarding. Proceedings of the CHI 2002: ACM Conference on Human Factors in Computing Systems, CHI Letters 4(1). ACM, Minneapolis, MN, pp. 17–24.

Zhai, S., Conversy, S., Beaudouin-Lafon, M., Guiard, Y., 2003. Human on-line response to target expansion. Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI Letters 5(1). ACM, Fort Lauderdale, FL, pp. 177–184.

Zhai, S., Kong, J., Ren, X., 2004. Speed-accuracy trade-off in Fitts' law tasks—on the equivalency of actual and nominal pointing precision. International Journal of Human–Computer Studies.