5-1-2021

Speed-Accuracy Trade-Offs and General Systems Performance Theory: Novel Application to Fitts' Law and Beyond

Lawrence R. DiSalvi

George V. Kondraske

SPEED-ACCURACY TRADE-OFFS AND GENERAL SYSTEMS PERFORMANCE THEORY: NOVEL APPLICATION TO FITTS' LAW AND BEYOND

Lawrence R. DiSalvi and George V. Kondraske
University of Texas at Arlington
Arlington, Texas USA

Speed-accuracy trade-offs have long been of interest in human performance. General Systems Performance Theory (GSPT) was motivated by human performance measurement and modeling needs. It has subsequently been applied in those and other areas. In GSPT, all system performance attributes are modeled using a resource construct. Systems are characterized by multi-dimensional performance capacity envelopes (PCEs). The systems of interest here are considered to possess limited speed and accuracy performance resources defining a two-dimensional PCE. When considering human movement, relevance to Fitts' law was conjectured. In multiple Fitts' paradigm tasks, we found a near-perfect correlation between Index of Performance (IP) and PCE area. An almost exact prediction was obtained when scaled by Fitts' index of task difficulty (ID). While the well-known Fitts' law equation does not contain accuracy explicitly, the GSPT-derived expression ($CP_{S\text{-}A}$= ID • Speed • Accuracy) contains both speed (motions/s) and accuracy (hits/attempts). Concepts are applicable beyond human movement; e.g., to visual, auditory, or other information processing types.

The notion of speed-accuracy trade-off has garnered attention wherever human performance is important, including aviation. One definition (Zimmerman, 2011) describes it as "the complex relationship between an individual's willingness to respond slowly and make relatively fewer errors compared to their willingness to respond quickly and make relatively more errors." Motivated by human performance measurement and modeling needs, Kondraske introduced General Systems Performance Theory (GSPT) and has applied it to problems in those areas (e.g., Kondraske, 2011). In GSPT, systems are characterized by multi-dimensional performance capacity envelopes (PCEs) with each dimension representing different performance resource types (e.g., speed, accuracy, etc.). Whereas Zimmerman's definition suggests a behavioral perspective, GSPT provides a performance viewpoint.

In human performance, systems that accomplish human movement represent one type of interest. In this context, despite a plethora of discussions and debates regarding various details, Fitts' law (Fitts, 1954) is widely recognized as that which explains speed-accuracy trade-offs. This is indeed interesting given that it does not explicitly contain speed or accuracy parameters.

For decades, our group has used Fitts' alternating tapping between two targets paradigm in various human performance measurement systems (Kondraske et al., 1984; Kondraske, 1990; Kondraske and Stewart, 2008; Saganis et al., 2020), relying on his *Index of Performance* (IP) with accuracy adjustment (see below) for scoring. When GSPT emerged, it inspired new thinking about many interesting research topics which we have since pursued on a somewhat opportunistic and case-by-case basis as time permits. Among those was an insightful moment regarding composite scores for movement tasks used in our human performance tests. Specifically, GSPT teaches that the mathematical product of speed and accuracy (i.e., the volume of a 2-dimensional PCE) would be a conceptually sound and meaningful composite. Moreover, it was conjectured that a composite so formed might, or should, have some relationship to Fitts' IP. A strong correlation was subsequently confirmed (Kondraske, 1999). While intriguing, other challenges dominated our research agenda until recently when we explored this in greater detail.

Background

Fitts' Law

Fitts's law is an empirical model considered to explain speed-accuracy tradeoffs in human movement with origins in Shannon's information theory and the concept of channel capacity. Fitts' early experiments focused on worker efficiency-related pointing motions in production line and assembly tasks that intrinsically involve speed and accuracy. A key element of Fitts' work is the definition of an *Index of Difficulty* (ID) for such tasks, where ID is a function of motion size or amplitude (A) and the target width (W). While readers are directed elsewhere for details, rationale behind ID, and variations of the initial idea (Fitts, 1954; MacKenzie, 1992), we use a form that has been argued to have desirable characteristics (Sourkoroff and MacKenzie, 2004):

$$\text{ID} = \log_2 (\text{A/W} + 1) \tag{1}$$

ID has units of bits or, more specifically, bits/motion.

Basic Fitts' Law studies usually employ an alternating upper extremity task (e.g., left-to-right, right-to-left) with multiple motions per trial, measurement of movement time (MT, units = s/motion), and sets of trials that exercise A and W over ranges of interest. To identify where participants work near their capacity limit, trials with too many errors as well as those with none or "near none" are excluded. Generally, researchers follow Fitts' approach, selecting trials with close to a four percent error rate and the corresponding MTs. Assuming a Gaussian distribution of landings at the target region, this corresponds to having target width boundaries at the ±2 standard deviation unit points.

With the above elements in place, Fitts' law is stated in what has been termed its usual form (MacKenzie, 1992) as:

$$\text{MT} = a + b \cdot \text{ID} \tag{2}$$

where a and b are coefficients determined by linear regression (i.e., best fit to MT and ID data). It states that MT varies linearly with ID. The intercept a is generally a small adjustment. Fitts dubbed the inverse of the slope b the *Index of Performance* (IP) with units of bits/s. When a is zero or not explicitly considered, as was the case in Fitts' original paper:

$$\text{IP} = (1/\text{MT}) \cdot \text{ID} = (1/\text{MT}) \cdot \log_2 (\text{A/W} + 1) \tag{3}$$

Explicit speed and accuracy terms are not present in either Equation (2) or Equation (3).

Not surprisingly in retrospect, an "adjustment for accuracy" was proposed by Crossman in 1957 in an unpublished report. The method involves the notion of an *effective target width* (W_e). That is, for a repetitive motion dataset with an actual target width (W) where the error rate is not constrained to 4%, determine the target width (W_e) that effectively yields a 4% error rate. While several accounts exist, MacKenzie (1992) provides a concise description of the somewhat cumbersome calculation for the case where only the error rate (*Error*) is known:

$$W_e = W \cdot 2.066/z(1 - Error/2) \;\text{ for } Error > 0.0049\%, \;\; W_e = W \cdot 0.5089 \text{ otherwise} \quad (4)$$

The term $z(\alpha)$, the inverse cumulative distribution function, returns the z-score where the area under the curve is α%. Considering both bell curve tails, a 4% error rate would require the evaluation of $z(1 - 0.04/2) = z(0.98) = 2.0537$ and $W = 1.00 \cdot W_e$. For *Error* >4%, $W_e > W$.

General Systems Performance Theory

Various aspects of GSPT and their rationale are described elsewhere (e.g., Kondraske, 2011). Briefly, GSPT's objectives are to provide: 1) a conceptual basis to define and measure *all* aspects of *any* system's performance; 2) a conceptual basis to analyze any task and facilitates system-task interface assessments; and 3) identification of the principles that explain success/failure in any given system-task interface. In GSPT, all aspects of a system's performance capacity are modeled with a resource construct. Each *performance resource* represents one dimension of a multidimensional performance space and the goal of system characterization from a performance perspective is to determine its *performance capacity envelope* (PCE). GSPT teaches to expect and how to define a PCE for any system. Another key feature is the *nonlinear, threshold effect* associated with resource economic mathematics at play in system-task interfaces; i.e., performance resource availability must exceed demand ($R_A \geq R_D$) for "success" give rise to new methods of performance prediction in complex tasks.

GSPT and Fitts' Law

As noted, GSPT suggested that speed-accuracy PCEs be considered for human movement systems. For any PCE defined according to GSPT, a single number composite performance measure (i.e., $CP_{S\text{-}A}$) can be obtained as the PCE volume (or, in this case, area):

$$CP_{S\text{-}A} = k \cdot \text{Speed} \cdot \text{Accuracy} \quad (5)$$

where *k* is a scaling constant. It was unavoidable but to wonder how this metric would relate to Fitts' IP, which we had been incorporating extensively in the design of instruments to measure aspects of human coordination (e.g., Kondraske, 1990; Saganis et al., 2020).

Methods

Three experiments using de-identified data previously collected during research and development of human performance measurement tools were conducted. Each involves a version of the alternating tapping Fitts' paradigm. Data for Experiments I and II was collected to evaluate a modular human performance measurement system (Kondraske, Potvin, Tourtellotte, and Syndulko 1984; Kondraske, 1990) as part of a center grant. Participants (n = 452; 267 female, 185 male), self-declared healthy, ranged in age from 7 to 83 years (mean = 36.4, sd = 16.6). Many contributed more than one test session. For Experiment III, data was obtained from a dataset created with a web-based tool (RC21X) for cognitive and neuromotor performance measurement (Saganis et al., 2020). Measures from 3rd and 4th self-administered sessions were used for participants (n= 33; 3 female and 30 male) ranging in age from 10 to 74 years (mean = 48.5, sd = 15.8). When asked if healthy, 19 responded “yes” while 14 responded “no”.

Experiment I data was collected with a computer-based device with six touch sensor regions (two targets with A = 40.6 cm and W = 1.6 cm, each flanked by two large error regions) during an upper extremity task requiring medial-lateral reciprocal motion. Experiment II employed a similar device (A = 52.0 cm, W = 10.5 cm) in a lower extremity task, while Experiment III data was collected using RC21X in an upper extremity test involving reciprocal tapping between the “A” and “L” keys on a computer keyboard. All set-ups involved the execution of two 10s trials. For Experiments I and II, the better of two trials was retained and available for use and many participants contributed more than one test session. For Experiment III, data from both trials was used.

For each trial (745, 745, and 66 for Experiments I, II, and III respectively), measures of movement speed (i.e., 1/MT) and accuracy (%) were used to compute the accuracy adjusted (Equation(4)) Fitts’ IP and the GSPT-based composite CP_{S-A} using Equation (5) with k = ID for each A/W case (W = *actual width*). Scatter plots were prepared to explore relationships.

Results

Figure 1 facilitates comparison of Fitts’ IP and GSPT-based CP_{S-A}. Pearson’s r ranged from 0.96 to 0.99. The average of the absolute value of the percent difference (i.e., average of $|100 \cdot (CP_{S-A} - IP)/IP|$) ranged from 2.6% for Experiment II to 8.1% for Experiment I.

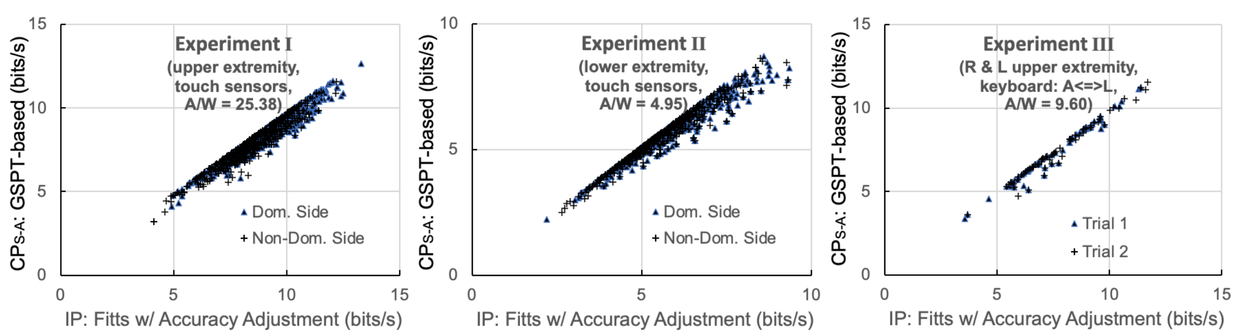

Figure 1. In three contexts and over a wide range of values, strong agreement was found between the simple-to-compute GSPT-based Composite Performance (CP_{S-A} = ID • Speed • Accuracy) and Fitts’ Index of Performance (IP) with accuracy adjustment.

Discussion

In multiple Fitts’ repetitive-motion, fixed-target-width pointing tasks with a range of ID and IP values, we found a very strong, near-perfect correlation between Fitts' Index of Performance (IP, bits/s) and GSPT-based Composite Performance (CP_{S-A}), with close agreement of actual values when *k* in Equation (5) is Fitts' ID computed using the *actual target width*. While we could have used simulated data, we opted to argue by using real experimental data to define the ranges of interest and perhaps communicate this interesting finding in a more direct and powerful way.

The simple and intuitively attractive Equation (5), based on the generalized concept of a PCE and its volume, provides essentially the same result as the relatively more complex and awkward Fitts' IP with the accuracy adjustment expression. There are clearly *some* differences. Preliminary analysis shows a relationship between these differences and task accuracy (or error) rate, with the largest differences for large error rates (e.g., 50%). It is not feasible, at present, to characterize such differences as "error", as that would require the assumption that the IP value is indeed a solid gold standard. The extensive Fitts'-related literature questioning aspects of conceptualization and proposing various tweaks, in part, argues against that premise.

We have noted that the well-known expressions of Fitts law and IP do not explicitly contain speed or accuracy variables. However, we also note that 1/MT has the units of speed. In CP_{S-A}, speed is expressed as motions/s. Comparing expressions for IP with the accuracy adjustment and CP_{S-A} leads to a focus on the equivalency of [Accuracy (%) • $\log_2$ (A/W + 1)] and [$\log_2$ (A/W_e + 1)]. Data presented here illustrates a very high equivalency level. Expounding on this via both conceptual and mathematical avenues is likely.

Equation (5) contains a performance index (CP_{S-A}), speed, accuracy, and task index of difficulty (ID). Given any three, the fourth can be computed. Of course, there have been similar interests in the use of Fitts' law. One can argue they contribute to the existence of an "accuracy adjustment" to allow consideration of arbitrary accuracy rates (i.e., not just 96%). With such motives, Wobbrock and colleagues (2008) proposed what they termed an "error model for pointing based on Fitts' law". With Accuracy(%) defined as 1 - Error(%), our preliminary review suggests that an "accuracy model for pointing", based on GSPT and Fitts' law, will provide similar results with a simpler expression (i.e., Accuracy (%) = CP_{S-A}/(ID • Speed), where CP_{S-A} is equivalent to Fitts' IP). Further analysis is warranted.

Fitts' contribution with regard to the definition of ID is not only useful but elegant in its simplicity. MacKenzie (1992) discusses this type of appeal with regard to Fitts' law. While an apparently sound conceptual basis can be argued for the accuracy adjustment, there are some initial assumptions involved that lead to the computational complexity present in the adjustment and a detraction from the simplicity appeal. Our results suggest a review of such assumptions and their impact in defining the speed-accuracy tradeoff in human pointing motions.

The powerful idea of PCEs can be traced to an aerospace context, where the dimensions of performance (i.e., speed, altitude, and range) and the metrics used (i.e., a larger value means "more" of that quantity) *naturally* lead to a PCE. This is not the case or so clear for many systems, where the commonly employed metrics (e.g., error vs. accuracy; time as a *speed-related* measure) do not result in an envelope! One might wonder about his modeling efforts if Fitts incorporated the notion of a speed-accuracy PCE. We emphasize that the speed-accuracy PCE can apply to not only human movement, but also information processing in general. It is perhaps unfortunate that work in those areas relies on time and error measures instead speed and accuracy, subverting identification of a PCE. It is also useful to observe that PCEs of greater dimensionality that incorporate dimensions of performance other than speed and accuracy are likely to be of great interest in human performance characterizations and modeling efforts.

References

Fitts, P. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391.

Kondraske, G. V. (1990). A PC-based performance measurement laboratory system. *Journal of Clinical Engineering*, 15(6), 467-478.

Kondraske, G. V. (1999). Determination of Fitts' Index of Performance using constructs of General Systems Performance Theory and implications for motor control performance capacity measurement. In A.M. Sherwood (Ed), *Seventeenth Annual Houston Conference on Biomedical Engineering Research*, Houston (p 86).

Kondraske, G.V. (2011). General Systems Performance Theory and its application to understanding complex system performance. *Information • Knowledge • Systems Management*, 10 (1-4), 235-259.

Kondraske, G. V., Potvin, A. R., Tourtellotte, W. W., & Syndulko, K. (1984). A computer-based system for automated quantitation of neurologic function. *IEEE Transactions on Biomedical Engineering*, 31(5), 401-414.

Kondraske, G. V., & Stewart, R. M. (2008). Web-based evaluation of Parkinson's Disease subjects: objective performance capacity measurements and subjective characterization profiles. In *Proceedings of the 30th Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Vancouver, 799-802.

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

Saganis, X., Goreczny, A. J., Kondraske, G., Berman, M., Comnick, I., Allen, T., & Nussbaum, P. D. (2020) Test–retest reliability of RC21X: a web-based cognitive and neuromotor performance measurement tool. *Archives of Clinical Neuropsychology*, 00 (2020), 1–7. doi.org/10.1093/arclin/acaa056

Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *Int. J. Human-Computer Studies*, 61 (2004), 751–789. doi:10.1016/j.ijhcs.2004.09.001

Wobbrock, J. O., Cutrell, E., Harada, S., & MacKenzie, I. S. (2008). An Error Model for Pointing Based on Fitts' Law. In *CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1613–1622. doi.org/10.1145/1357054.1357306

Zimmerman, M.E. (2011) Speed–Accuracy Tradeoff. In: J.S. Kreutzer, J. DeLuca, B. Caplan (Eds) *Encyclopedia of Clinical Neuropsychology*. Springer, New York, NY. doi.org/10.1007/978-0-387-79948-3_12