Time-Accuracy Data Analysis: Separating Stimulus-limited and Post-stimulus Processes

Ángel Cabrera
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332-0170
phone: (404) 853 0192
angel@psy.gatech.edu

Tony J. Simon
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332-0170
phone: (404) 894 2681
tony.simon@psych.gatech.edu

Abstract

Time-accuracy functions are obtained by measuring the accuracy of a subject's responses at various levels of stimulus presentation time. Unlike reaction time (RT) measurements, which convey information about the entire set of processes taking place between the onset of the stimulus and the production of a response, time-accuracy functions (TAFs) focus on a subset of those processes, namely stimulus-limited processes. Stimulus-limited processes are responsible for the extraction of the perceptual information that is necessary for the elaboration of a response. Post-stimulus processes take care of selecting and executing the response based on the information extracted by stimulus-limited processes. This paper presents a method of analysis that allows us to (a) extract estimates of the duration and variance of stimulus-limited processes from individual TAFs and (b) combine these estimates with RT data in order to induce the duration and variance of post-stimulus processes. The method is illustrated with data from a subitizing (speeded enumeration) task.

Introduction

The relationship between speed and accuracy of responding conveys information about the dynamics of mental processing at a level of detail beyond the scope of reaction time measurements (Wickelgren, 1977; Meyer, Irwin, Osman & Kounios, 1988). Consider for instance the three hypothetical speed-accuracy curves in Figure 1. Curves like these represent the relationship between the time available to the subject to deliberate a response and the average accuracy of the responses in different experimental situations. In most cases, accuracy starts out at a low (chance) level for small values of time and then increases monotonically as more time becomes available. In the asymptotic region, accuracy no longer depends on time, but on factors such as task difficulty, individual capacity, or amount of noise in the stimuli (Norman and Bobrow, 1975).

Imagine that the three curves in Figure 1 are the true speed-accuracy functions of three cognitive processes an experimenter is trying to differentiate. The three curves show some important differences. Performance in task A increases gradually with time, whereas, in tasks B and C, it changes dramatically from chance to asymptote within a few milliseconds. In addition, accuracy in tasks A and B reaches a higher asymptote than it does in C.

What would the experimenter find by running a typical RT task? By simply looking at the figure, it seems that all three tasks take about the same amount of time to reach asymptote. If subjects are instructed to emphasize accuracy, it is likely that average RT of correct trials would be somewhere around t_1 ms in all three cases, and therefore, no significant differences among the tasks would be found. Further analysis of the subject's errors would probably reveal a difference between task C and tasks A and B, but A and B might still remain indistinguishable.

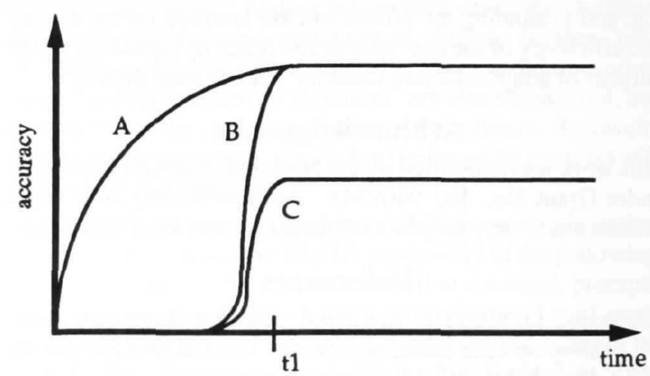

Figure 1. Hypothetical time-accuracy functions

Speed-accuracy tradeoff research

How then can we assess the correspondence between speed and accuracy in order to capture those differences that seem to escape RT data? Several methods have been used in the past to induce subjects to trade accuracy for speed (Wickelgren, 1977): verbal instructions, differential reward of speed and accuracy, external signals that indicate when a response must be produced, etc. Speed-accuracy curves can also be estimated by partitioning all the responses (correct or incorrect) into uniform RT intervals and then computing the percent of correct responses within each interval. The strengths and limitations of each of these methods have been discussed elsewhere (e.g. Wickelgren, 1977). In any case, no matter what method is used, speed-accuracy designs treat both speed and accuracy as *dependent* measures.

Time-accuracy functions

A different way of dealing with speed-accuracy dependencies involves measuring accuracy as a function of the time during which the stimulus is directly available to the subject (Lohman, 1986; Loftus, Duncan & Gehrig, 1992; Kliegl, Mayr & Krampe, 1994; Simon, Cabrera & Kliegl, 1993). Kliegl and his colleagues refer to this presentation time manipulation method as the "cognitive psychophysics" approach for its resemblance to the general psychophysical paradigm in which performance is analyzed as a function of the available amount of some relevant resource (Norman and Bobrow, 1975). Here the relevant resource is the time during which the subject has access to the perceptual information needed to produce a response, and performance is the subject's accuracy in that task.

The difference between this approach and more traditional speed-accuracy research is more important than it may initially appear. In time-accuracy research "presentation time is under experimental control, whereas in speed-accuracy tradeoff research response latency is a dependent variable" (Kliegl, Mayr & Krampe, 1994, p. 136). This contrast is not only methodological: it affects the kinds of substantive inferences that can be drawn from the results. Speed-accuracy tradeoff data, like RTs, convey information about the entire sequence of processes taking place from the onset of the stimulus until the production of the response. In contrast, the shape of a time-accuracy function obtained through the psycophysical method is only descriptive of a subset of those processes, namely *those subprocesses that rely on the physical availability of the stimulus.*

Stimulus-limited vs. post-stimulus processing

Like Salthouse (1981), we will refer to the processes that require direct access to the stimulus as "stimulus-limited". In addition, we will define "post-stimulus" processes as those processes that rely on the result of stimulus-limited processes and that, consequently, do not operate until stimulus-limited processes end (spontaneously or otherwise). We have suggested that the transition region of a time-accuracy function reflects dynamic characteristics of stimulus-limited processing. We also know that RTs convey information about the entire set of processes including both stimulus-limited and post-stimulus stages. If we can somehow extract information from the TAFs in a form that is directly comparable with RT information, we could infer the properties of post-stimulus processing by subtraction *à la* Donders (1868) or through more sophisticated deconvolution techniques (Smith, 1990).

For this stage decomposition to be possible, we must assume that (a) the two stages are connected in series, (b) the duration of each of the two stages is a random variable with certain probability distribution, and (c) that the duration of the two stages is statistically independent from each other. These assumptions are common to most of the stage decomposition techniques used in cognitive psychology (Sternberg, 1969; Salthouse, 1981). Although the seriality and independence assumptions have been the center of a great deal of controversy (see Townsend & Ashby, 1983), there is a significant body of findings that has shown the concept of processing stage to be useful in understanding human cognition (Salthouse, 1981). Resolving this debate is beyond the scope of this paper. Rather, we are making these assumptions from a systems identification standpoint. The usefulness of the assumptions will be indicated by the kinds of empirical distinctions they ultimately lead to. As we will show later (see also Simon & Cabrera, 1995) some of our own data in the domain of subitizing indicates that the stimulus-limited / post-stimulus distinction may capture important conceptual differences.

Analysis of time-accuracy functions

Time-accuracy functions have been analyzed in the past by fitting mathematical functions (typically negatively accelerated exponential curves) to the data. The functions so obtained are then compared with one another (between or within subjects) in terms of either the parameters resulting from the curve fitting procedure (Kliegl, Mayr & Krampe, 1994; Simon, Cabrera & Kliegl, 1993) or the predicted presentation time for some prespecified level of accuracy (Wickelgren, 1977; Lohman, 1986).

These parametric techniques have several limitations. First, the choice of a particular mathematical model may be hard to justify theoretically. More often than not, models are justified in a post-hoc fashion based on goodness-of-fit measures and not on *a priori* theoretical grounds (c.f. Kliegl, Mayr and Krampe, 1994). Second, the resulting parameters are difficult to interpret outside the specific family of mathematical functions being considered, and consequently, they can not be compared to RT data or even to parameters from other mathematical functions.

How can we extract information about stimulus-limited processing from a time-accuracy function (TAF) in a form that can later be used in comparisons with RT data? There are at least two main ways in which time-accuracy functions can be interpreted (Wickelgren, 1977; Meyer et al., 1988). One of them consists of assuming a decision process that gradually and stochastically builds a response tendency starting as soon as the stimulus is presented. When presentation time is experimentally restricted, subjects are forced to produce a premature response based on incomplete information. This translates into the typical time-accuracy patterns in which accuracy increases monotonically and asymptotically with presentation time. Under this approach then, TAFs represent diagrams of information (or response activation) accumulation (e.g. McClelland, 1979).

The second interpretation (the one that will be adopted here) assumes an underlying discrete, all-or-none process with a randomly distributed termination time. In other words, it assumes that there is a minimum amount of presentation time T necessary for a stimulus to be processed successfully, and that this amount of time is a random variable with a certain distribution. If the presentation time for a given trial t is less than T, the stimulus can not be processed, and a random response is produced. However, if t is greater than T, the stimulus can be processed successfully and the accuracy of the response only depends on the capacity limitations of the subject or the noise in the stimuli. Saying that T is a random variable amounts to saying that the exact duration of the underlying all-or-none process is not fixed across trials. This means that any given value of presentation time can be too short in some trials and yet long enough in other trials --which will lead to random performance in some trials and to asymptotic performance in other trials.

The all-or-none interpretation of stimulus-limited processes might initially seem counterintuitive. Even though, based on the data alone, both interpretations are equally valid (see Meyer et al., 1988), it can be argued that some amount of partial information can be extracted even when presentation time is too short for a fully confident response to be possible. We have two reasons to assume the all-or-none interpretation. The first reason is an analytical one and will hopefully become clear in the next section. The second reason is an empirical one: this assumption leads to results that could be critical in understanding the phenomenon of subitizing. Whether this empirical advantage will generalize to other areas of cognition or not, remains an open question.

Difference Time-Accuracy Functions

The analysis problem under this interpretation can be stated as: "given $A(t)$, the observed accuracy for each value of presentation time t, obtain an estimate of $F(t)$, the cumulative probability distribution of the underlying random variable T." Let us refer to the asymptotic level of $A(t)$ as A_{max}, and to the chance level as A_{rand}. Since $F(t) = P(t \leq T)$, the observed time-accuracy function $A(t)$ will be related to $F(t)$ according to:

$$A(t) = F(t)A_{\max} + [1 - F(t)]A_{rand}, \quad (1)$$

In other words the observed accuracy will depend on the probability of t being above or below T, times the level of accuracy reached in each case (A_{max} and A_{rand}). Given $A(t)$ we can then obtain $F(t)$ as

$$F(t) = \frac{A(t) - A_{rand}}{A_{\max} - A_{rand}} \quad (2)$$

Once we have an estimate $F(t)$ of the cumulative probability distribution of the underlying T, it is straightforward to derive T's probability density function $f(t)$ as the derivative of $F(t)$, $f(t) = \frac{dF(t)}{dt}$. In practice, we may not have an analytical expression for $F(t)$ but a set of samples for different values of time $F(t_i)$. In this case, we can obtain the probability function $p(t_i)$ through simple subtraction:

$$p(t_i) = F\left(t_i + \frac{\Delta t}{2}\right) - F\left(t_i - \frac{\Delta t}{2}\right) \quad (3)$$

where Δt represents the size of the interval between consecutive samples. The functions $f(t)$ or $p(t_i)$ so obtained will be referred to as Differential Time-Accuracy Functions (DTAFs). Figure 2 shows the DTAF (marked as •-•), obtained from a hypothetical discrete logistic TAF (marked as *- -*).

By interpreting DTAFs as estimates of probability distributions, we have access to the whole spectrum of stochastic techniques normally used in analyzing RT distributions (Townsend & Ashby, 1983). The most immediate computations DTAFs can address are estimates of central tendency and variance of stimulus-limited termination time. For instance, we can obtain the mathematical expectation or mean of T as

$$\mu_T = \sum_i t_i p(t_i) \quad (4)$$

and its variance as

$$\sigma_T^2 = \sum_i (t_i - \mu_T)^2 p(t_i) \quad (5)$$

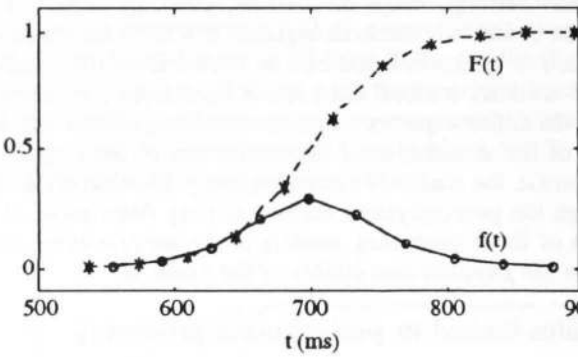

Figure 2. Logistic TAF and its corresponding DTAF.

The *median* of T is the point of intersection of $F(t)$ with the .5 constant line. Computing the median from the DTAF is equivalent to using psychophysics' typical 50% criterion, which estimates the amount of resource necessary to complete a task as the value of resource for which performance reaches 50% of the asymptote. The *mode* of T corresponds to the peak of $f(t)$, which in turn corresponds to the inflation point of $F(t)$. Although in the hypothetical curve in Figure 1, mean, median and mode are the same, this is not necessarily so for other distributions. The median and the mode tend to be more robust to the presence of noise and outliers in the data than the mean, but the mean offers two important advantages: (a) it is the maximum-likelihood estimate of T, and (b) stage duration means are additive under the seriality and independence assumptions.

For DTAF analysis to be feasible, the original TAFs must be monotonic and non-decreasing (otherwise, we could obtain distributions with negative values of probability!) What this constraint really means is that accuracy must increase or remain constant as time increases. Fortunately, this is the case with most tasks of interest in cognitive science (Norman & Bobrow, 1975). In real situations, even when monotonicity is to be expected, errors of measurement may induce small violations. It is the experimenter's responsibility to decide, based on theoretical grounds, whether to consider these irregularities as random errors of measurement or as reflection of some unexpected underlying pattern. In the first case, the irregularities can be smoothed out prior to further analysis through curve fitting, convolution or any other method.

Post-stimulus processing

The most important advantage of the DTAF procedure is that it produces results that are given in terms of process duration or termination time. This has the key advantage of allowing results from time-accuracy designs to be compared and combined with duration estimates obtained through other methods. In particular, they can be combined with RT data, in order to estimate the time characteristics of post-stimulus processes. Remember that we defined stimulus-limited and post-stimulus processes as two independent and serial stages. It can be shown (e.g. Townsend & Ashby, 1983) that the distribution of the duration of the serial combination of two independent stages can be obtained as the mathematical convolution of the distributions of the two stages. Furthermore, the mean duration of the combined process is equal to the sum of the mean duration of the individual stages, and so is the variance.

The additivity property of serial stages allows us to interpret mean RT as the sum of the mean duration of stimulus-limited processes plus the mean duration of post-stimulus processes, and the same thing for the variances. In other words, we can estimate the mean duration and variance of post-stimulus processes as the difference between the mean and variance of the observed RTs and the estimates of mean and variance of stimulus-limited processes obtained through DTAF analysis. The fact that the estimates of mean and variance for the whole process are obtained very differently than the mean and variance of the stimulus-limited subprocesses does not preclude us from being able to combine and compare them with one another because they represent the same underlying construct: process duration.

Subitizing: A case study

To illustrate the whole procedure, we have selected some data from one subject in one of the experiments we are carrying out in our on-going research on subitizing (Simon, Cabrera & Kliegl, 1993; Simon & Cabrera, 1995). Subitizing refers to the ability of humans to identify the numerosity of small sets of objects rapidly an accurately. This phenomenon is important for its implications in the understanding of visual attention (Trick & Pylyshyn, 1994) as well as the development of basic numeric ability (Geary, 1995). According to Trick and Pylyshyn's meta-analysis (1994), enumerating 1 to 4 objects takes an average of 40 to 120 ms per item, whereas enumerating more than 5 objects requires between 250 and 370 ms per item.

As we discuss elsewhere in this volume (Simon & Cabrera, 1995), all the different attempts to explain the phenomenon of subitizing split the process of number judgments into two separate stages, although they vary in the nature of the subprocesses attributed to each of the stages. Roughly speaking, the first stage would be responsible for the perception and encoding of the stimulus while the second stage would be responsible for response choice and execution. Depending on the exact roles that are attributed to each of the two stages we can make different predictions about how the number of objects in a display should affect the duration of each of the stages. Unfortunately, the RT data that has been collected so far is not powerful enough to assess the effect of numerosity on each of the two processing stages.

Simon, Cabrera & Kliegl (1993) reported some results from a time-accuracy study that showed a discontinuity between small and large number of objects very similar to the one found in RT studies. Our goal now is to assess the effects of numerosity on stimulus-limited and post-stimulus stages. Our preliminary results show that numerosity mainly affects stimulus-limited processes within the subitizing range but affects both stimulus-limited and post-stimulus for larger numerosities (Simon & Cabrera, 1995). Here is an illustration of how the method is actually applied.

Subjects are presented with a row of letters "o" in the center of a computer screen for a period of time that is varied from trial to trial. Subjects are asked to enumerate the letters in the display and to produce a verbal response as fast as they can that is timed by using a voice-activated relay. Response times as well as accuracy of the responses are recorded for every trial. Throughout the whole session, subjects are presented with rows of 2 to 8 items at ten different levels of presentation time 20 times each. Average accuracy for each numerosity at each of the ten levels of presentation time is used to build the TAFs (one per numerosity). The ten levels of presentation time are not the same for all numerosities. They are actually selected based on pilot data so that accuracy ranges from chance at the lowest presentation times to perfect at the largest. Figure 3 shows seven TAFs obtained from one subject for numerosities 2 to 8 under presentation times varying overall from 35 to 2500 ms. Each of these seven curves are our raw $A(t_i)$'s.

Notice how almost every $A(t_i)$ shows some small violations of monotonicity even though there is a clear increasing trend of accuracy with time. There are several methods to smooth out those irregularities from the raw $A(t_i)$'s. These include exponential or logistic curve fitting and kernel filtering techniques. To avoid having to make the assumptions those techniques rely on, we opted for a very simple algorithm that flattens out any irregular segment to the average of the neighboring points (Cabrera, 1994). These adjusted versions of the $A(t_i)$s are then rescaled according to Equation 2 to obtain estimates of the $F(t_i)$. In the case of subitizing, most subjects reach perfect accuracy provided enough presentation time, so we considered A_{max} = 1.0. Since our subjects were informed that numerosity varied between 2 and 8, we assumed that A_{rand} = 1/7. Figure 4 shows the effect of smoothing and rescaling one of the raw TAFs.

From each $F(t_i)$ we compute a $p(t_i)$ (DTAF) by subtraction of consecutive values of $F(t_i)$ (Eq. 3) and then extract the mean and variance of stimulus-limited processes from the resulting DTAFs (Eqs. 4 & 5). Figure 5 shows the resulting means with their corresponding confidence intervals ($\pm\sigma_T$) as a function of numerosity. By subtracting these values from the means and variances of the RT data, we obtain the means and variances of post-stimulus processes (Figure 6). The data from this one subject illustrate how numerosity affects stimulus-limited and post-stimulus processes differently (see Simon & Cabrera, 1995, for further details).

Conclusions

Reaction time data convey information about the entire set of processes taking place from the onset of the stimulus until the production of a response. The relationship between accuracy of responses and the amount of time during which the stimulus is physically available to the subject conveys information about the subset of processes responsible for the extraction of information from the stimulus, namely stimulus-limited processes. By combining RT and time-accuracy data we can then estimate some of the characteristics of the remaining processes (post-stimulus processes). We have shown how the stimulus-limited / post-stimulus distinction can help us understand the nature of cognitive processes at a level of detail that escapes more traditional RT designs. In particular, it has helped us identify the processing stages that are involved in the estimation of numerosity, a problem with interesting implications in visual attention and basic mathematical ability (Simon & Cabrera, 1995).

One of the advantages of this method is the fact that it does not rely on specific parametric assumptions about the form of the data. Experimental manipulations often have effects on the shape of TAFs, and these effects may make it hard for a simple mathematical model to fit all the data across conditions. In our earlier subitizing studies (Simon, Cabrera & Kliegl, 1993) a two parameter exponential model achieved very good fits for numerosities up to 5 but very poor fits thereafter, which forced us to discard portions of data that would otherwise be perfectly valid. DTAF analysis does not make specific assumptions about the shape of the data and can therefore be applied no matter what the shape of the TAF is (with the only constraint of monotonicity).

A second strength of DTAF analysis is its applicability to individual data. Several researchers (Siegler, 1987; Ashby, Maddox & Lee, 1994) have emphasized the need to extract as much information as possible at the individual level.

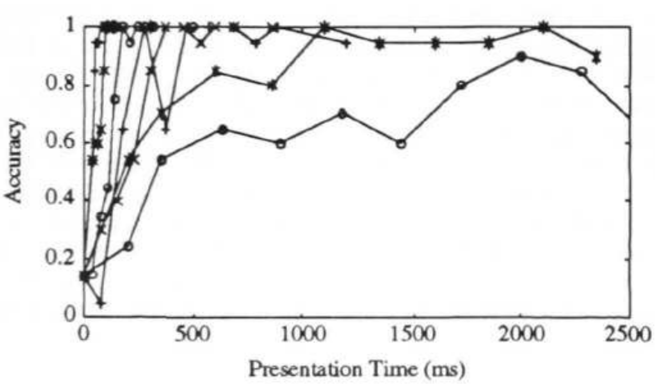

Figure 3. Original TAFs for numerosities 2 to 8.

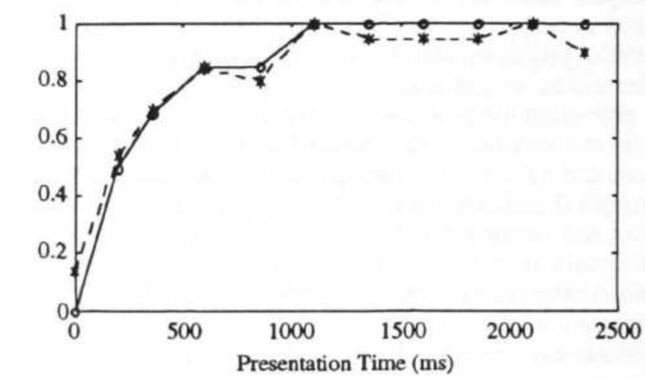

Figure 4. Raw $A(t_i)$ (--*--) for 7 items and its corresponding $F(t_i)$ (-o-) after smoothing and rescaling to the [0, 1] interval.

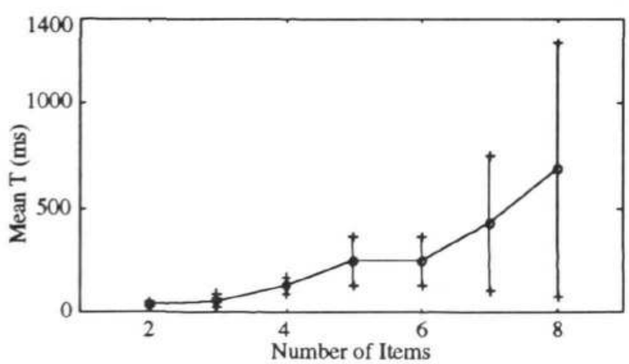

Figure 5. Duration of stimulus-limited subprocesses as a function of numerosity.

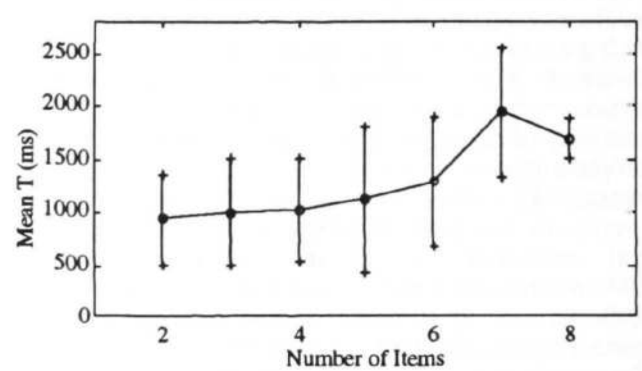

Figure 6. Duration of post-stimulus subprocesses as a function of numerosity.

As information theorists well know, any kind of data manipulation can only eliminate information, or, at the very best, keep it constant. If relevant information about individual processing is filtered out through premature data aggregation, it is unlikely that further analyses will be able to recover it. At some point of course, aggregation is necessary in order to generate meaningful conclusions from the data. But it is crucial to devote as many resources as possible to the early stages of analysis and to defer aggregation across subjects to the final stages. The DTAF method is particularly well suited for this.

Finally, the methodology presented here links TAF paradigms with all the stochastic techniques that, in the past, were exclusive to response latency data (Townsend & Ashby, 1983). Here we saw how statistical estimates obtained from TAF and RT distributions can be combined in order to address questions about the internal structure of cognitive processes. Convolutional analysis, statistical dominance and other stochastic constructs are perfectly applicable to DTAFs in the same way as they are used in the context of RT distributions.

Wickelgren (1977) argued that theories of information processing could no longer ignore speed-accuracy tradeoff data, yet general theories of cognition still rely fundamentally on response latencies to assess their capacity to account for human performance (e.g. Newell, 1990). Response latencies yield a single point of a speed-accuracy or time-accuracy function. Many important processing differences that are reflected in the patterns of dependency between time and accuracy can not be detected by RT designs. Speed-accuracy data in general, and time-accuracy functions in particular, raise new challenges for cognitive theorists and model builders.

Acknowledgments

The first author was funded by a MEC/Fulbright Scholarship from the Spanish Ministry of Education and Science and the U. S. Information Agency. Throughout this research we have received invaluable insights from Chris Hertzog, Reinhold Kliegl and Tim Salthouse. Our kindest thanks to Christa Dell, Alan Kersten, Jonah Lunken and Sandeep Vaishnavi for their collaboration in this project.

References

Ashby, F., Maddox, W. & Lee, W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. Psychological Science, 5, 144-151.

Cabrera, A. (1994). Controlling time and measuring time: A tow-stage model of cognitive processing of visual stimuli. Cognitive Science Report Series, GIT-CS-1994/34. Georgia Institute of Technology.

Donders, F. C. (1909). On the speed of mental processes. (English translation of original 1868 article.) Acta Psychologica, 30, 412-431.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. American Psychologist, 50, 24-37.

Kliegl, R., Mayr, U. & Krampe, R. T. (1994). Time-accuracy functions for determining process and person differences: An application to cognitive aging. Cognitive Psychology, 26, 134-164.

Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time-course of perceptual information that results from a brief visual presentation. Journal of Experimental Psychology: Human Perception and Performance, 18, 530-549.

Lohman, D. F. (1986). The effect of speed-accuracy tradeoff on sex differences in mental rotation. Perception and Psychophysics, 39, 427-436.

McClelland, J. L. (1979). On the time relations of mental processes: an examination of systems of processes in cascade. Psychological Review, 86, 287-330.

Meyer, D. E., Irwin, D. E., Osman, A. M. & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. Psychological Review, 95, 183-237.

Newell, A. (1990). Unified Theories of Cognition. Cambridge, MA: Harvard University Press.

Norman, D. A. & Bobrow, D. (1975). On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.

Salthouse, T. A. (1981). Converging evidence for information-processing stages: A comparative-influence stage-analysis method. Acta Psychologica, 47, 39-61.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. Journal of Experimental Psychology: General, 116, 250-264.

Simon, T. J., Cabrera, A. & Kliegl, R. (1993). A new approach to the study of subitizing as distinct enumeration processing. In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society, Hillsdale, NJ: Lawrence Erlbaum.

Simon, T., & Cabrera, A. (1995). Evidence for subitizing as a stimulus-limited phenomenon. Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society.. Hillsdale, NJ: Lawrence Erlbaum.

Smith, P. L. (1990). Obtaining meaningful results from Fourier deconvolution of reaction time data. Psychological Bulletin, 108, 533-550.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. Acta Psychologica, 30, 276-315.

Townsend, J. T. & Ashby, F. G. (1983). The Stochastic Modeling of Elementary Psychological Processes. New York: Cambridge University Press.

Trick, L. & Pylyshyn, Z. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. Psychological Review, 101, 80-102.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information dynamics. Acta Psychologica, 41, 67-85.