Proceedings of Meetings on Acoustics

Volume 26 http://acousticalsociety.org/

171st Meeting of the Acoustical Society of America

Salt Lake City, Utah

23-27 May 2016

Psychological and Physiological Acoustics: Paper 3aPP3

Unified analysis of accuracy and reaction times via models of decision making

Samuel R. Mathias
Department of Psychiatry, Yale University, New Haven, CT; samuel.mathias@yale.edu

Over the past half century, cognitive scientists have developed sophisticated models to explain decision making in simple forced-choice experiments. These models are able to explain both accuracy and reaction times (RTs). On a practical level, these models deal with many of the nuisances that come with analyzing RT data, such skewed RT distributions, and speed-accuracy trade-offs. On a theoretical level, they quantify the dissociable elements of the decision-making process, providing a clearer picture of the effects of experimental manipulations and individual differences. However, despite being extremely popular in the cognitive sciences, such models are almost never used in psychoacoustics. This paper briefly introduces the most widely used decision-making model (the drift-diffusion model), and provides some recommendations for its use in future psychoacoustical research.

Published by the Acoustical Society of America

1. INTRODUCTION

Accuracy and reaction times (RTs) share almost equal standing as dependent variables within the cognitive sciences. However, as illustrated in Fig. 1, this is not true in psychoacoustics: far fewer articles published in the Psychological and Physiological Acoustics section of *The Journal of the Acoustical Society of America* since 2010 contained the words "reaction time(s)" or "response time(s)" ($n = 12$) than contained the words "accuracy," "sensitivity," or "thresholds" ($n = 216$) in the title or abstract. By contrast, the same analysis performed with the *Journal of Experimental Psychology: Human Perception and Performance* yielded more similar numbers of articles reporting accuracy ($n = 116$) and RTs ($n = 83$).

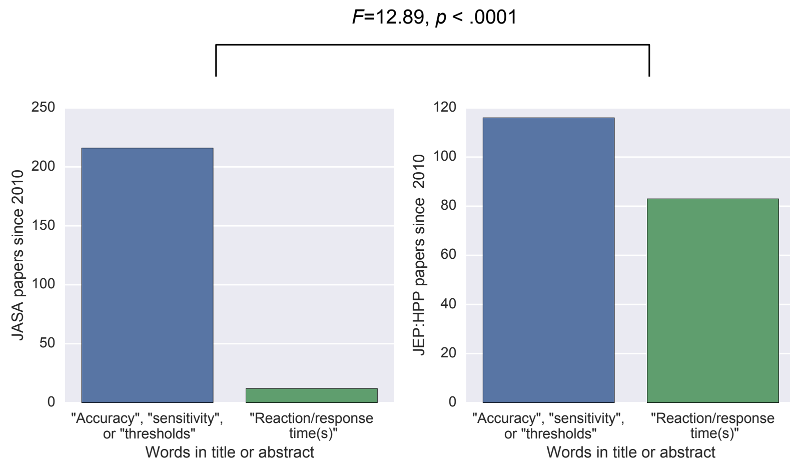

***Figure 1.* Results of a crude literature search. Fisher's exact test shows that the proportions of articles reporting RTs are significantly different between the two journals.**

What could explain the underrepresentation of RTs in psychoacoustics? One possibility is that there are a number of potential stumbling blocks in the analysis of RT data. For example, RT distributions tend to exhibit a strong positive skew, which vitiates the choice of an appropriate measure of central tendency, and calls into question the validity of conventional parametric statistical tests, such as the ANOVA. Another stumbling block is the phenomenon of the speed-accuracy tradeoff, or the non-independence of RTs and task difficulty (see Heitz, 2014). Thirdly, RTs can be heavily influenced by myriad factors that are not under experimental control and may not be particularly interesting to the researchers, such as hardware-related latencies, sensory encoding time, and sluggishness in motor preparation and execution.

Fortunately, researchers in other areas have already encountered these issues, and have largely dealt with them by developing models of decision making. These models fit to both accuracy and RT data. On a practical level, they provide a way of dealing with many of the nuisances inherent in RT data: they generally predict positively skewed RT distributions, they naturally deal with speed-accuracy trade-offs, and they allow researchers to remove much unwanted variation from the data. On a theoretical level, they quantify the dissociable elements of the decision-making process, providing a clearer picture of the effects of experimental manipulations and individual differences. However, despite their popularity elsewhere, these models are almost never used by psychacousticians.

This paper provides a brief introduction to the drift-diffusion model (DDM), the most popular extant model of decision making. It describes how the DDM can be considered an extension of statistical hypothesis testing, how to interpret its parameters, and provides practical recommendations for fitting the model to real data. This introduction is intended to be simple, not thoroughly didactic, and readers are pointed to in-depth articles where appropriate.

2. THE DDM

A. LOGIC OF THE MODEL

The DDM, as with many models of decision making (see Bogacz *et al.*, 2006; Gold & Shadlen, 2007), is normally only applied to experiments that require subjects to choose between two response alternatives on each trial (i.e., 2AFC). Although the DDM can be modified to explain data from other experimental paradigms (e.g., Gomez *et al.*, 2009), it is not straightforward to do so, and is beyond the scope of this paper (but see the "Concluding remarks").

The DDM can be seen as an extension of the sequential probability ratio test (SPRT), a statistical method of hypothesis testing (Wald, 1945), to human decision making (Bogacz *et al.*, 2006). Consider a sequential stream of evidence, $X = \{x_0, x_1 \cdots x_n\}$, generated by either of two competing hypotheses, H_0 and H_1. The first step in the SPRT is to set two thresholds (or "boundaries"), a and b, one for each hypothesis. Then, as each new piece of evidence is accumulated, the running log-likelihood ratio is computed:

$$\mathrm{LLR}_t = \sum_{i=0}^{t} \ln \left[{P(x_i|H_o)}\big/{P(x_i|H_1)} \right],$$

where t indexes the step, and $P(\cdot)$ is the likelihood, or probability of the evidence under one of the hypotheses. This procedure continues so long as LLR_t falls between a and b, but when one of the boundaries is crossed, the corresponding hypothesis is chosen and the process is terminated. Under the SPRT, LLR_t can be considered to go on a random walk until either boundary is crossed. The continuous analog of the random walk, suitable for modeling real RT data, is the Wiener diffusion process, which forms the basis of the DDM (Fig. 2).

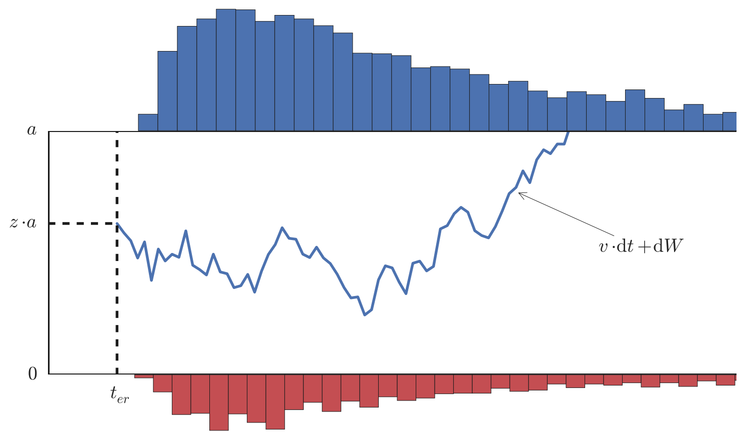

Figure 2. **Illustration of the Wiener diffusion process in the context of the DDM. The middle panel shows the "decision space," and an example instance of the process, with the decision variable as its ordinates and time as its abscissae. The blue and red histograms are RT data corresponding to the two response alternatives.**

Instances of the Wiener process are generated by the stochastic differential equation:

$$\mathrm{d}y = v \cdot \mathrm{d}t + \mathrm{d}W,$$

where y is the decision variable, t is the elapsed time, v is the "drift rate," and W is Gaussian noise. In terms of the SPRT, the drift rate is the average outcome of each sub-test, $\ln[P(x_i|H_o)/P(x_i|H_1)]$. Without loss of generality, b is fixed at 0, which allows a to reflect the separation between the boundaries. The figure contains two other parameters that are not part of the SPRT, but are necessary to produce realistic RT distributions: z is the relative starting point of diffusion process, where a value of 0.5 is equidistant between the boundaries, and t_{er} (called "non-decision time") is essentially a constant that introduces a lag to the RT distributions.

The four free parameters v, a, z, and t_{er} make up the "basic" DDM. However, the basic DDM does not usually provide an adequate quantitative fit to real RT data, and must be augmented with up to four additional parameters, forming the "full" DDM (see Wagenmakers, 2009). First, normally distributed variability in drift rate, s_v, is introduced to account for when error responses are systematically slower than correct responses (Ratcliff, 1978). Second, uniformly distributed trial-to-trial variability in starting point, s_z, is introduced to account for faster error responses (Laming, 1968). Third, uniformly distributed variability in non-decision time, $s_{t_{\mathrm{er}}}$, is introduced to account for gradual rises in the leading edge of the RT distribution (Ratcliff & Tuerlinckx, 2002). Fourth, a proportion of the data, p, are sometimes assumed to come from an admixture of uniformly distributed "contaminant" responses (Ratcliff & Tuerlinckx, 2002).

B. INTERPRETING THE PARAMETERS

Because the DDM is rooted in statistical hypothesis testing, each of its free parameters has a clear role in the decision-making process. The drift rate v can be considered to represent the quality of the evidence and is closely analogous to sensitivity (d') from signal detection theory: large v means that decisions are reached easily, and when $v = 0$, the competing alternatives are indistinguishable. The boundary separation a can be considered a measure of response caution: when a is small, decisions require a small amount of accumulated evidence and are made quickly, but when a is large, decisions require more evidence and take longer. The starting point z can be considered a measure of *a priori* response bias, and non-decision time t_{er} reflects other factors that influence RTs but are not related to the decision process *per se*, such as motor preparation. The trial-to-trial parameters s_v, s_z, and $s_{t_{\mathrm{er}}}$ reflect just that—variability across trials in the instantaneous values of v, z, and t_{er}. Finally, p reflects the proportion of responses not generated according to the DDM, such as ones where the subject experienced a lapse in attention.

Extensive prior research has shown that these interpretations are generally valid. For example, Voss *et al.* (2004) showed that for a simple color-discrimination task, manipulations to stimulus discriminability, speed-accuracy instructions, payoff structure, and ease of executing a motor response yielded the expected changes in drift rate, boundary separation, starting point, and non-decision time, respectively. Moreover, when the DDM is fitted to data from a recognition-memory experiment, s_v is greater for old stimuli than new stimuli, which tallies with the unequal-variance signal detection account of recognition memory (Starns & Ratcliff, 2014). For more examples, see Wagenmakers (2009) and Ratcliff and McKoon (2008).

The DDM allows researchers to draw conclusions that would not be possible by considering accuracy and RT data separately, some particularly good examples of which can be found in the literature on practice effects. Practice could improve efficiency in extracting information from the stimulus ("perceptual learning"), influence the placement response boundaries, and/or simply speed up motor responses. Using the DDM, these three effects can be isolated and estimated separately, and the differential effects of task instructions on these different kinds of practice effects can be evaluated (e.g., Dutilh *et al.*, 2009; Zhang & Rowe, 2014).

C. FITTING THE DDM

Unfortunately, fitting the DDM can be rather challenging, and is best accomplished via specialist software. Currently, the best choices are RWiener (Wabersich & Vandekerckhove, 2014b), an R package (R Foundation for Statistical Computing, https://www.R-project.org); JAGS-Wiener (Wabersich & Vandekerckhove, 2014a), a module for JAGS (Martyn Plummer, http://mcmc-jags.sourceforge.net); and HDDM (Wiecki *et al.*, 2013), a Python package (Python Software Foundation, https://www.python.org/). Since none of these are stand-alone software, you will need to first install the appropriate base software (JAGS, R, or Python), plus any other required dependencies.

Two of the three software packages listed above (JAGS-Wiener and HDDM) implement *Bayesian hierarchical* DDMs. Here, "Bayesian" refers to Bayesian inference, a method of statistical inference that relies on Bayes' theorem (see Gelman *et al.*, 2004), which is sharply contrasted with frequentist inference, the more traditional method of inference that underlies most common statistical tests, such as the ANOVA. "Hierarchical" refers to the fact that parameters are estimated hierarchically (see Gelman & Hill, 2007); that is, it is assumed that for a given subject, the value of a given parameter (e.g., v) is a random variable drawn from a group-level distribution, and the parameters of this distribution (e.g., μ_v and σ_v) are estimated together with the subject-level parameters. In general, Bayesian hierarchical models have better parameter-recovery properties than their frequentist non-hierarchical counterparts, and this is particularly true for the DDM, since prohibitively large numbers of trials are often needed to obtain reasonable parameter estimates using traditional fitting methods (see Vandekerckhove *et al.*, 2011; Wiecki *et al.*, 2013). Bayesian inference holds many other advantages over frequentist inference (see Gelman *et al.*, 2004). However, it is important to point out that, if one chooses to use either JAGS-Wiener or HDDM, it is not appropriate to apply subsequent frequentist statistical tests to the recovered parameters; all further analysis should be done within the Bayesian framework.

3. CONCLUDING REMARKS

Models of decision making allow the unified analysis of accuracy and RT data. They have both considerable practical advantages, related to dealing with nuisances inherent in RTs, and considerable theoretical ones, allowing insights into decision making that would not be possible by other means. It is a shame that such models are not more popular in psychoacoustics, since they make good use of a dependent variable that is regularly thrown away in the majority of psychoacoustical studies.

Given its close relationship to statistical theory, empirical validity of its parameters, and the availably of fitting software, the DDM is the obvious choice of decision-making model for 2AFC data. For situations that involve more than two alternatives per trial, the linear ballistic accumulator (LBA) model may be a better choice (Brown & Heathcote, 2008). The LBA model is fundamentally different to the DDM in terms of its underlying assumptions, but there is an approximate correspondence between LBA and DDM parameters (Donkin *et al.*, 2011). Although the LBA model is not rooted in statistical decision making and is less popular than the DDM, it has the considerable advantage of being much simpler than the DDM, which makes it easier to fit, and makes it easier to generalize to experiments with any number of alternatives (for a tutorial on fitting the LBA, see Donkin *et al.*, 2009).

An exciting possible direction for future research is to explore the relationships between decision-making model parameters and physiological measurements. For example, a recent study found that trial-to-trial variations in low-frequency oscillatory activity in the subthalamic nucleus (measured with EEG) predicted the values of boundary separation from the DDM in a visual coherent-motion discrimination task (Herz *et al.*, 2016; see also Frank *et al.*, 2015). In psychoacoustics, there is growing interest in the ability of physiological measurements of peripheral auditory function to predict performance in hearing tasks (e.g., Bharadwaj *et al.*, 2015). Models of decision making may provide a way to more specifically characterize the nature of any such relationships. Consider, for example, the frequency-following response (FFR), a putative index of the quality of peripheral temporal coding: one might hypothesize that trial-by-trial measurements of the FFR would correlate with drift rate during a frequency discrimination task.

REFERENCES

Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S., and Shinn-Cunningham, B. G. (**2015**). “Individual differences reveal correlates of hidden hearing deficits,” J. Neurosci. **35**, 2161-2172

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., and Cohen, J. D. (**2006**). "The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks," Psychol. Rev. **113**, 700-765.

Brown, S., and Heathcote, A. J. (**2008**). "The simplest complete model of choice reaction time: Linear ballistic accumulation," Cogn. Psychol. 57, 153-178.

Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., and Wagenmakers, E.-J. (**2009**). "A diffusion model decomposition of the practice effect," Psychon. Bull. Rev, **16**, 1026-1036.

Donkin, C., Averell, L., Brown, S., and Heathcote, A. (**2009**). "Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator," Behav. Res. Methods **41**, 1095-1110.

Donkin, C., Brown, S., Heathcote, A., and Wagenmakers, E.-J. (**2011**). "Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes?" Psychon. Bull. Rev. **18**, 61-69.

Frank, M. J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J. F., and Badre, D. (**2015**). "fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning," J. Neurosci. **35**, 485-494.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (**2004**). *Bayesian data analysis* (Chapman & Hall/CRC Press, New York).

Gelman, A., and Hill, J. (**2007**). *Data analysis using regression and multilevel/hierarchical models* (Cambridge University Press, Cambridge, UK).

Gold, J. I., and Shadlen, M. N. (**2007**). "The neural basis of decision making," Annu. Rev. Neurosci. **30**, 535-574.

Gomez, P., Ratcliff, R., and Perea, M. (**2009**). "A model of the go/no-go task," J. Exp. Psychol. Gen. **136**, 389-413.

Heitz, R. P. (**2014**). "The speed-accuracy tradeoff: History, physiology, methodology, and behavior," Front. Neurosci. **8**, 150.

Herz, D. M., Zavala, B. A., Bogacz, R., and Brown, P. (**2016**). "Neural correlates of decision thresholds in the human subthalamic nucleus," Curr. Biol. **26**, 916-920.

Laming, D. R. J. (**1968**). *Information theory of choice-reaction times* (Academic Press, London, UK).

Ratcliff, R. (**1978**). "A theory of memory retrieval," Psychol. Rev., **85**, 59-108.

Ratcliff, R., and McKoon, G. (**2008**). "The diffusion decision model: Theory and data for two-choice decision tasks," Neural Comput. **20**, 873-922.

Ratcliff, R., and Tuerlinckx, F. (**2002**). "Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability," Psychon. Bull. Rev. **9**, 438-481.

Starns, J. J., and Ratcliff, R. (**2014**). "Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis," J. Mem. Lang. **70**, 36-52.

Vandekerckhove, J., Tuerlinckx, F., and Lee, M. D. (**2011**). "Hierarchical diffusion models for two-choice response times," Psychol. Methods **16**, 44-62.

Voss, A., Rothermund, K., and Voss, J. (**2004**). "Interpreting the parameters of the diffusion model: An empirical validation," Mem. Cogn. **32**, 1206-1220.

Wabersich, D., and Vandekerckhove, J. (**2014a**). "Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example)," Behav. Res. **46**, 15-28.

Wabersich, D., and Vandekerckhove, J. (**2014b**). "The RWiener package: An R package providing distribution functions for the Wiener diffusion model" The R Journal **6**, 49-56.

Wagenmakers, E.-J. (**2009**). "Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy," Eur. J. Cogn. Psychol. **21**, 641-671.

Wald, A. (**1945**). "Sequential tests of statistical hypotheses," Ann. Math. Statist. **16**, 117-186.

Wiecki, T. V., Sofer, I., and Frank, M. J. (**2013**). "HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python," Front. Neuroinform. **7**, 14.

Zhang, J., and Rowe, J. B. (**2014**). "Dissociable mechanisms of speed-accuracy tradeoff during visual perceptual learning are revealed by a hierarchical drift-diffusion model," Front. Neurosci. **8**, 69.