Psychological Bulletin
1995, Vol. 118, No. 2, 285-299

0033-2909/95/$3.00

The Application of Fourier Deconvolution to Reaction Time Data: A Cautionary Note

Ching-Fan Sheu
DePaul University

Roger Ratcliff
Northwestern University

The Fourier transform method in conjunction with frequency domain smoothing techniques has been suggested as a powerful tool for examining components in a serial, additive reaction time model (P. L. Smith, 1990). Robustness and sensitivity to violations of the assumptions of serial model of this method are evaluated. When an incorrect distribution was used in recovering an unobserved component, results gave no information to show that an incorrect distribution was used, and the results were just as interpretable as those obtained using the correct distribution. These results demonstrate that the assumptions underlying the method cannot be assessed by the result of deconvolution, and the method cannot show that the purported component is actually from the serial combination.

The Fourier transform method has a long and distinguished history in many areas of science (physics, chemistry, engineering, biology, etc.), and it has been used in psychology in the domain of reaction time (as well as in computational vision and audition). A recent evaluation of the Fourier deconvolution method was carried out by Smith (1990). He examined possible reasons for a relative lack of success in previous applications of the method (e.g., Green & Luce, 1971; Kohfeld, Santee, & Wallace, 1981) and presented specific ways to overcome the problems. In particular, he argued that the reason for failure is because variability or noise in the observed reaction time distribution is amplified by Fourier decomposition. He presented the methods for reducing the effect of such variability and concluded that the Fourier transform method in conjunction with noise reduction methods (filtering) can be a powerful and accurate tool for examining components in a serial additive reaction time model.

In many experiments, participants were asked to make decisions, and the dependent variable was the time taken to make those decisions. The total time for a decision was assumed to be the sum of the times taken by a series of subprocesses required to reach the decision. It is sometimes theoretically important to obtain estimates of the distributions of times taken by the separate subprocesses. For example, Green and Luce (1971) estimated the decision component of processing from a theoretical model and extracted the residual component from the reaction time distribution. The residual component was then compared across experimental conditions to test if it was invariant as the theory predicted. (Parameters of the explicit model of the decision process could be estimated from the tails of the observed distributions; see Green & Luce, 1971.) In another study of simple reaction time to an auditory stimulus, Burbeck and Luce (1982) used the reaction time distribution at a very high auditory signal intensity as an estimate of the residual distribution assuming that very little decision time was involved when the stimulus was intense. They used this residual to deconvolve observed distributions from signals at lower intensities, then used the extracted estimates to test the fit of a specific model of the decision times.

In this article, we examine how the distribution of one subprocess can be obtained from the distribution of another subprocess and the distribution of the total response time for the decision (when the two subprocesses are serially organized). We investigate one method of extracting the distribution of a subprocess, a method that uses Fourier deconvolution. With this method, if we have the reaction time distribution of the combination of two subprocesses (i.e., their convolution) as well as the reaction time distribution of one of them, then Fourier deconvolution can be used to extract an estimate of the distribution of the other subprocess (in practice, one subcomponent is assumed to be a combination of processes, e.g., a residual reaction time component includes encoding and response output processes). Because many models of human information processing either assume or predict response times at the distribution level, the potential use of the Fourier deconvolution method is very broad. If the time distribution of an underlying process can be successfully extracted, then the investigators can compare not only the means, medians, or modes but also the distributional shapes in evaluating competing process models.

A major problem that has bedeviled reaction time research

Ching-Fan Sheu, Department of Psychology, DePaul University; Roger Ratcliff, Department of Psychology, Northwestern University.

This research was supported by National Institute of Mental Health Grants MH 44640 and MHK00871 and National Institute on Deafness and other Communication Disorders Grant R01-DC01240. We are indebted to Geoffrey J. Iverson and Mark Perkins for their helpful discussions on the topic of Fourier transform.

We would like to thank Ehtibar Dzhafarov for raising the issues discussed in the Appendix and for comments on this article. We would like to thank Mathieu Koppen, Gail McKoon, Neil Mulligan, Sue O'Curry, Philip Smith, and James Townsend for comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Ching-Fan Sheu, Department of Psychology, DePaul University, 2219 North Kenmore Avenue, Chicago, Illinois 60614-3504. Electronic mail may be sent via Internet to csheu@sheu.psy.depaul.edu.

is the predictions of one model mimicking the predictions of another. Although one model of a process (e.g., a serial model) might fit the data well, other models (e.g., a parallel model) might fit equally well. This makes discriminating between models difficult, especially when limited data are available (see, Ratcliff, 1988; Townsend, 1972, 1990; Townsend & Ashby, 1983; Van Zandt & Ratcliff, 1995). The issue of mimicking has been examined in many domains of reaction time modeling, and the general conclusion has been that it is often difficult, if not impossible, to work backward from a small set of reaction time data (including distributions of reaction times) to uncover the architecture responsible for the data. The problem is that the same set of results can be obtained from a number of different architectures (Van Zandt & Ratcliff, 1995).

The aim of this article is to extend evaluation of the Fourier deconvolution method by examining the degree to which it is subject to mimicking problems. If an incorrect choice for one subprocess distribution is made and the resulting distribution of the second subprocess appears reasonable, is there any way to know that the choice was incorrect? For example, Green and Luce (1971) found that the recovered densities of the residual latencies had negative loops, which indicated that the assumptions of the model on which they based the deconvolution were suspect. In this article, we investigate whether and under what circumstances incorrect choices of distributions for deconvolution can be detected.

Fourier Deconvolution

Take Donders's classical idea that the total observed reaction time is the sum of a number of successive times (Donders, 1969/1868–1869) and consider the case in which there are only two such times: the decision latency D and the residual latency R. It is natural to think of these latencies as distributions because the time observed in an experimental task is not always the same. In general, it is also assumed that the two latency distributions D and R are independent. We focus on the simple reaction time structure of additive and stochastically independent components because the structure forms the basis of much modeling of reaction times in psychology (see, e.g., Burbeck & Luce, 1982; Green & Luce, 1971; Kohfeld et al., 1981; Luce, 1986). However, there can be many different reaction time structures. For example, Dzhafarov (1992) compared the simple structure of sums of independent components with additive decompositions into a structure with perfectly positive interrelated components. Rouder (1993) discussed a nonasymptotic, distribution-free decomposition of simple reaction time. Roberts and Sternberg (1993) analyzed, in a more elaborate version, Dzhafarov and Schweickert (1994) structures that contain components combined by operations other than summation.

A stochastic model explains how long it takes to complete a task by predicting the shape of the time distributions. To be useful, a theoretical time distribution must be able to fit the reaction time measurements obtained from experiments. The decision time of a task is usually not directly observable. However, the residual time may be estimated from a simple, isolated part of the same task. In a discrete serial model, testing the distributional predictions of a component time becomes possible if the time distribution of such a component can be extracted from the measurable total time and can be treated as data.

The method that has gained the most popularity in determining the distribution of a time component in the above setting is the Fourier deconvolution technique (Burbeck & Luce, 1982; Green & Luce, 1971; Smith, 1990). It is also widely used to address the problem of restoring acoustic signals in audition and reconstructing images degraded by blurs in vision. Mathematically, the probability density function of the sum of the two reaction times D and R is the convolution of the density functions, f_D and f_R. Conveniently, the convolution of f_D and f_R corresponds to ordinary multiplication of their Fourier transforms, ϕ_D and ϕ_R. Thus, one method used to convolve two distributions is to take Fourier transforms and multiply them to find the inverse transformation of the product. This is sometimes much easier than solving the convolution integrals directly, especially because there are extensive tables available of inverse transformations (e.g., Oberhettinger, 1973). Therefore, from the relation (and provided that $\phi_R \neq 0$),

$$\phi_D = \frac{\phi_T}{\phi_R}, \tag{1}$$

where ϕ_T is the transform of T, we obtain the density of D by inverting the transform ϕ_D. In symbols,

$$f_D(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\zeta t} \frac{\phi_T(\zeta)}{\phi_R(\zeta)} d\zeta. \tag{2}$$

In the method used in psychology (Green, 1971; Green & Luce, 1971), the transform and its inversion are based on the fast Fourier transform (FFT) algorithm (Brigham, 1988). Because the FFT routine is efficient and widely available, the FFT deconvolution method has practical advantages over other methods of extracting a component time process.

Green (1971) and Smith (1990) discussed the Fourier transform technique and its applications to reaction time data extensively. Smith applied the method to real data (with time samples in the thousands) and found the results satisfactory. It was suggested by Smith that the method is a powerful tool for examining distribution of a component time in a serial, additive model. In this article, we examine issues that raise potential problems and limitations for the application of the method. They are the performance of the method in the presence of noise (or bad data) and the problem of mimicking. It turns out that the method is reasonably robust against noisy data. However, the problem of mimicking is serious. We show that if incorrect (but reasonable, i.e., right-skewed) distributional assumptions are made about one of the components of the convolution, then the deconvolved results are still interpretable, and the method gives few if any clues that the assumptions were wrong. Many investigators (e.g., Luce, 1986; Ratcliff, 1988; Townsend, 1972, 1990) have examined mimicking problems in reaction time research. The consequences of mimicking in model testing, however, have not been addressed in the discussion of deconvolution, most likely because the major focus has been to make the method a usable tool. Nevertheless, we should not forget that the FFT method does not validate the additive and independent assumptions under which it is licensed to op-

erate. Neither can it resolve on its own the different models of reaction times. The major advantage of the method actually lies in its efficiency as a data-processing tool and its independence to prior model assumptions (Bendat & Piersol, 1971). Until more sophisticated diagnostic techniques are developed for deconvolution, the use of the method in research should primarily be for exploratory analysis.

Noise and Mimicking

In this section, we examine the sources of noise in the reaction time data and show why mimicking is inherent in the deconvolution method.

In most experimental situations, the data collected are not free from uncontrollable measurement errors or noise. This implies that the densities f_T and f_R are only known to within some experimental error. Under such circumstances, the solution to the convolution integral

$$f_T(t) = \int_{-\infty}^{\infty} f_D(t - x) f_R(x) dx \qquad (3)$$

cannot be obtained in the manner shown in Equation 2 (Turchin, Kozlov, & Malkevickz, 1971). The solution may be neither unique nor stable. To counter it, we can do our best to ensure that spurious responses do not contaminate our data. Real data are discrete samples of finite values, and on a digital computer the transforms in Equation 2 have to be implemented by their discrete counterparts. Naturally, the quality of the output depends on the sampling rate and the sampling range of the input. In the signal processing literature, there is extensive discussion on how to obtain reasonable deconvolution results using the discrete Fourier transform (see, e.g., Silverman & Pearson, 1973). Most methods have relied on filtering to improve the quality of the results because the presence of the noise has reduced the power of the signal. In particular, noise corrupts the accuracies of the high-frequency components in the FFT estimates of the densities, which, in turn, lead to large spurious, oscillatory components in the estimates of $f_D(t)$. To obtain better estimates, Green (1971) introduced the use of the Hamming window for smoothing the data. Smith (1990) emphasized the importance of conditioning the values of the transformed estimates through filtering in the frequency domain and presented comparative results for a number of different smoothing and filtering techniques.

The filtering techniques either remove or attenuate the contribution of the high-frequency samples to the recovered time densities. This produces a problem when one wants to fit a predicted time distribution to the deconvolved results because sufficiently close functions have sufficiently close Fourier transforms and vice versa. Provided that a sufficient number of low-frequency components are the same and the rest is smoothed away by the filtering, two different time densities may pass through the FFT deconvolution procedure and come out resembling one another quite closely. In other words, we encounter a mimicking problem created by the deconvolution method.[1]

Basically, mimicking is a problem of uniqueness that occurs in different forms. On the one hand, there is mimicking within the same family of distribution models. Consider, for example, a gamma distribution with parameters r and α, in which r is a positive integer and $\alpha > 0$ is a scale parameter, as the predicted reaction time distribution. For $r = 5$, the gamma density function can be either a convolution of a two-stage gamma density and a three-stage gamma density (with the same scale parameter) or a convolution of a four-stage gamma density and a one-stage gamma density. On the other hand, there can be a statistical mimicking problem between two different classes of models. Ratcliff (1988) showed that predictions of the one-boundary diffusion model (Ratcliff, 1978) pass the tests for pure insertion in a serial model (Ashby, 1982; Ashby & Townsend, 1980), even though the models are incompatible. The mimicking problem in the deconvolution method is one of resolution—the ability of the method to sort distinct input into distinct output.

The previous discussion warns us that, without strong empirical support for the assumption regarding the shape of reaction time distributions, the FFT deconvolution method cannot be used to obtain the true density function describing the unobservable decision time. At present, we do not know if the problems of statistical mimicking would disappear or appear in a different form if Fourier transformations were replaced by other techniques of deconvolution. Thus, the same cautionary remark would apply to two other methods for estimating unobserved components of a serial model: the spline method used by Bloxom (1979) and the linear system approach adopted by Kohfeld et al. (1981). It should, perhaps, be emphasized that our work in the present article addresses only the mimicking problem induced by the Fourier approximation in practice.

Because the accuracy of the deconvolved results cannot be determined without a theoretically "correct" distribution, we study the problem of mimicking by simulation. More specifically, we examine the outcome of deconvolving a combination of times with an "incorrect" distribution for one of the component times. The questions are (a) whether something interpretable could result from such manipulation, that is, whether the recovered distribution is interpretable as the "true" unobserved distribution; and (b) whether the result gives any indication that the assumptions may be wrong. Of course, the quality of the deconvolved estimates cannot be evaluated without studying the impact of noise on the method first. Most of the results regarding the problem of noise are found in Smith (1990). One exception is that he did not investigate how the size of the sample used to estimate the latencies will affect the outcomes of deconvolution. This is important because the experimenter may not always be able to collect thousands of measurements. Also undocumented is how poor (or good) the result will be if the density to be extracted is not smooth everywhere, such as a peaked exponential density. The interest is to learn about how the method behaves in a simple case. These two questions are investigated first in the simulation. In the next section, we discuss the implementation and evaluation of the deconvolution

[1] McCullagh (1994) showed, by an example, that two visibly distinct distributions can have almost identical moment-generating functions (transforms). Conceivably, a similar example involving Fourier transforms might also be found. In such a case, the mimicking problem, numerical or theoretical, has nothing to do with the deconvolution method.

method, followed by a synopsis of the two reaction time distributions.

FFT Deconvolution Method

The FFT deconvolution method consists of the following basic steps:

1. The reaction time densities are grouped into a finite number of time samples in the form of a histogram. Keeping the rate and the range of sampling the same, the residual densities are estimated similarly.
2. The histogram estimates of the total and the residual latencies are transformed into frequencies samples using an FFT subroutine.
3. The power spectrum (the relative amplitude of the individual frequency components) of the total densities is examined to determine a suitable cutoff frequency for filtering. Filtering is performed by multiplying each value of frequency samples of the transformed total densities by the corresponding value of the window function at the same frequency.
4. The Fourier transform of the decision time is obtained point by point by the complex division of the filtered transform of the total densities (numerator) by the transform of the residual density estimates (denominator).
5. The inverse FFT subroutine is used to obtain the estimated decision densities in the time domain.

A few brief comments on these basic steps are in order. Smith (1990) recommended that all numbers be calculated with double precision accuracy to prevent rounding errors from becoming a source of high-frequency noise. The sampling range should be sufficiently long to include enough tail regions of a density function. Depending on the particular FFT algorithm used, proper arrangement of the data string is required before it can be processed. The FFT programs usually require that the number of input points equal some power of 2. Also note that some routines produce results that are correct in waveform but differ from the expected results by a scaling constant in amplitude. In the complex division of Step 4, it is important to take precautionary programming steps to avoid the division-by-zero situation at some data points.

Choosing and implementing filters is a large and important topic in signal processing. A full discussion of digital filtering can be found in Hamming (1983). In our article, we selectively use three different filters: the parabolic window function, the Hanning filter (Brigham, 1988), and the rectangular filter. Their functional forms are listed in this order for ease of reference. The first two functions are commonly used filters with good properties (Brigham, 1988; Press, Flannery, Teukolsky, & Vetterling, 1992). The rectangular filter is chosen because it has quite different filter characteristics from the other two. Smith (1990) gave both time and frequency profiles of these functions,

$$g(\zeta) = \begin{cases} 1 - \left(\dfrac{\zeta}{\zeta_o}\right)^2, & |\zeta| \le \zeta_o \\ 0, & |\zeta| > \zeta_o, \end{cases} \tag{4}$$

$$g(\zeta) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{2}\cos\left(\dfrac{\pi\zeta}{\zeta_o}\right), & |\zeta| \le \zeta_o \\ 0, & |\zeta| > \zeta_o, \end{cases} \tag{5}$$

$$g(\zeta) = \begin{cases} 1, & |\zeta| \le \zeta_o \\ 0, & |\zeta| > \zeta_o, \end{cases} \tag{6}$$

between the zero component and ζ_o, the upper bound of the filter's pass band. Essentially, the first two functions smooth out all frequency components beyond the pass band and attenuate the low-frequency of components according to their position between the zero component and the pass band. The rectangular filter, however, admits all frequency components smaller than ζ_o and sets all higher frequency components to zero. The experimenter usually varies the value of the parameter ζ_o until a satisfactory result is obtained. A simple rule to choose the initial value of ζ_o is to observe that the power spectrum of the convolution density estimates falls steeply to a minimum before rising. The frequency at the first minimum is often a good choice (Shaffer, Shaughnessy, & Jones, 1984).

Evaluating the Deconvolution Results

As mentioned in previous discussion, the deconvolved estimates usually contain oscillatory components signifying negative densities that are not possible. Another common error is that the peak value of the density is often underestimated. It is also typical to have deconvolved density estimates of a time distribution not to add up to one, thus violating the definition of a probability density function. (In contrast, Bloxom's, 1979, constrained cubic spline method produces an estimate that is a proper density.) To circumvent these difficulties, Burbeck and Luce (1982) proposed a procedure that eliminates the negative weights of the deconvolved estimates and treats the remaining coefficients as resulting from a censored distribution. In this article, we take the deconvolved result as is and examine the fit (or lack of it) between the estimate and true density by two error measures: the integrated square error (ISE) and the total variation (TV). These two measures are chosen to facilitate the comparison of our results with Smith's (1990) results. By definition, ISE is the sum of all squared differences between the estimated and true densities over the entire sample multiplied by the sample interval. In symbols,

$$\Delta T \sum_{k=0}^{N-1} [\hat{f}(t_k) - f(t_k)]^2, \tag{7}$$

where $\hat{f}(t)$ denotes the filtered estimate of $f(t)$, ΔT is the sampling interval of time, and N is the number of time samples. A large ISE value indicates a poor match between the true and deconvolved densities. Keeping all other things equal, the ISE value decreases as the number of observations in a simulation increases. On the other hand, TV is defined as the sum of the absolute difference between successive terms of the estimated density over the entire sample. In symbols,

$$\sum_{k=1}^{N-1} |\hat{f}(t_k) - \hat{f}(t_{k-1})|. \tag{8}$$

A density estimate may be said to achieve a certain smoothness criterion when it has a small enough TV value (relative to the pertinent theoretical distribution). In other words, the criterion of smoothness depends on the sampling intervals, as well as the

type and variance of the distribution we wish to estimate. For example, holding the sampling interval and the variance constant, an estimate of an 8-stage gamma distribution, on average, should have a smaller TV value than a 2-stage gamma distribution (with the same scale parameter). The requirement of a small TV value guarantees that the deconvolved result looks like a density function generated by the usual reaction time models. A large TV value in deconvolution is often caused by the oscillatory components at the tail of the estimate. In some cases the TV might be small, but there may be systematic deviations of the deconvolved distribution from the target distribution. This would result in a large ISE.

Two Reaction Time Distributions

The family of gamma distributions is often postulated to be the latency component of a serial model (see, e.g., Luce, 1986, for discussions) for the following reasons: (a) A sum of independent, identically distributed exponential random variables is gamma distributed; (b) an exponential distribution has been argued to be a component process of serial stage models (Townsend & Ashby, 1983)[2]; and (c) the family of gamma densities is closed under convolution. The gamma density we consider has the following form:

$$f(t; r, \alpha) = \frac{\alpha^r t^{r-1} e^{-\alpha t}}{(r-1)!}, \quad t > 0, \quad \alpha > 0, \tag{9}$$

where r is a positive integer and α is a scale parameter. The densities vanish for negative values of t. Its characteristic function is given by

$$\phi(\zeta) = \left[1 - \left(\frac{i\zeta}{\alpha}\right)\right]^{-r} \tag{10}$$

The gamma distribution arises as the waiting time until the occurrence of the rth event when one observes a sequence of events occurring at the rate of α events per unit time in accordance with a Poisson distribution (Feller, 1966). However, different theoretical considerations can lead to different distributions as models for reaction times. In contrast to the serial model, Ratcliff (1978) argued that the inverse Gaussian, being the distribution of the first passage time derived from a one-boundary diffusion process, is a good model for decision latency. Another reason for considering the inverse Gaussian distribution is that its hazard function can account for the general qualitative shape—peaked followed by constant asymptote—of empirical hazard functions such as simple reaction times to the offset of a weak, pure tone masked by wide-band noise (Burbeck & Luce, 1982). In practice, with a judicious choice of parameter values, any reasonably skewed distribution can fit the reaction time data. We used both inverse Gaussian and gamma distributions as models of decision latency in our simulation studies. Next, we briefly summarize some useful facts about the inverse Gaussian distribution.

The form of an inverse Gaussian probability density function is

$$f(t; \mu, \lambda) = \left(\frac{\lambda}{2\pi t^3}\right)^{1/2} \exp\left(-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right), \quad t > 0, \quad \mu > 0, \quad \lambda > 0, \tag{11}$$

and the densities vanish for negative values of t. The parameter μ is the mean of the distribution, and λ is a scale parameter. The characteristic function is

$$\phi(\zeta) = \exp\left((2\lambda)^{1/2}\left[\left(\frac{\lambda}{2\mu^2}\right)^{1/2} - \left(\frac{\lambda}{2\mu^2} - i\zeta\right)^{1/2}\right]\right). \tag{12}$$

The maximum likelihood estimator (MLE) of μ is the sample mean $\hat{\mu}$. The variance of the distribution is μ^3/λ. An unbiased estimator of the variance is a modified MLE given by $[n/(n-1)](\hat{\mu}^3/\hat{\lambda})$. The MLE of the scale parameter λ is

$$n\left[\sum_{i=1}^{n} \frac{1}{X_i} - \frac{1}{\hat{\mu}}\right]^{-1},$$

where X_i for $i = 1, 2, \ldots, n$ is a random sample from the inverse Gaussian distribution. To generate random observations of an inverse Gaussian distribution, we used the method of transformation with multiple roots (Michael, Schucany, & Haas, 1976). The procedure is described in Chhikara and Folks (1989).

Simulation Studies

We report four simulation studies. The first simulation examines the effect of sample size on the goodness of fit between the reconstructed and the expected density. The second simulation assesses the quality of a recovered exponential density using two different filters and varying the cutoff frequency. The third and fourth simulations investigate the validity of the method as a tool for testing distributional predictions of stage models. We provide two scenarios in which the total density is deconvolved by a theoretically incompatible component density.

Simulation 1. The quality of several deconvolved gamma densities based on 5,000 trials has been documented by Smith (1990). He used five different filters at various cutoff frequencies. To complement his results, we evaluated the effect of sample size on the deconvolved estimates of inverse Gaussian densities. We ran our simulation on the basis of four sample sizes: 512, 1,028, 2,056, and 4,096. It is useful to know how much the results can be improved by increasing the size of the observations. The method is not practical if the number of observations required to ensure a satisfactory result exceeds what can be achieved in an experimental setting.

We assumed that the total latency distribution (numerator) is a convolution of an inverse Gaussian distribution with M = 300 ms and SD = 100 ms and a normal distribution with M = 300 ms and SD = 60 ms. The residual latency distribution (denominator) is independent and identically distributed as the normal component in the convolution. Thus, the deconvolved results can be checked against the expected density of an inverse Gaussian distribution with M = 300 ms and SD = 100 ms. At each sample size, random observations are obtained from the total and residual latency distributions. We then constructed histograms for each distribution by grouping and normalizing

[2] Townsend and Ashby (1983) also pointed out that an exponential stage can implicate an intercompletion time associated with an exponential parallel model.

the observations into 256 bins, each having a width of 10 ms. For simplicity, we used a histogram to estimate the true density because other density estimates do not seem to improve the results of deconvolution (Smith, 1990). Each pair of total and residual histograms was then processed by the FFT deconvolution technique described earlier. After some experimentation, we set the value of the upper bound of the parabolic filter's pass band at 15, which yields consistent results.

The deconvolution was repeated 25 times, and the mean and standard deviation of the ISE and TV values are calculated. Table 1 summarizes the results for each sample size. As expected, the overall quality of the fit improves as the sample size increases. The average ISE value (based on 4,096 trials) of the deconvolved inverse Gaussian density estimates is comparable to Smith's (1990) results of an 8-stage gamma density (based on 5,000 trials). Moreover, the average TV value of the deconvolved estimates is almost identical to the true TV value of the expected inverse Gaussian density. This similarity suggests that the extracted density is as smooth as the expected density. This is confirmed by Figure 1 in which examples of deconvolved densities are shown.

To evaluate the deconvolved estimates of the inverse Gaussian density, we list in Table 1 the values (in parentheses) of ISE and TV measures for the corresponding histogram estimates. Not only are the deconvolved density estimates smoother than the histogram estimates but they also have smaller ISE values. This is due to the filtering operation (which produces smoothing) in the deconvolution. The discrepancies become smaller as the histogram estimates become smoother (and more accurate) as the sample size increases. This can be observed by noting that an eightfold increase in the number of trials yields only about a twofold reduction in the ISE value of the deconvolved estimates. In contrast, there is about a sevenfold reduction of ISE in the histogram estimates.

Figure 1 shows deconvolved estimates at different sample sizes superimposed on the inverse Gaussian density. We note that around the peak the values of the density estimates are lower than those of the true density and that there are oscillatory components at both tails of the estimates. As the sample size increases, the match of the values around the peak gets better and the oscillations dampen. In the frequency domain, increasing sample size translates to more agreement between the estimated and the exact spectrum. The significant improvement comes at the initial portion (low-frequency components). The effect is easier to see in a plot of amplitude (power) against frequency sample. In Figure 2 we show a sample of power spectra of the total time densities obtained at each of the four sample sizes. Note the rapid decrease of magnitude in power with increasing frequency. Power spectra like Figure 2 are used to determine how many frequency components can be retained in the deconvolution procedure without bringing in a large increase in the TV value.

Table 1
Effect of Sample Size on the ISE and TV Characteristics of the Deconvolved Estimates

N	ISE	TV
512	.056 ± .036	7.984 ± 1.047
	(.191 ± .038)	(39.984 ± 3.999)
1,024	.037 ± .017	7.034 ± 0.484
	(.098 ± .019)	(30.258 ± 3.052)
2,048	.032 ± .013	7.004 ± 0.365
	(.054 ± .010)	(22.273 ± 1.986)
4,096	.024 ± .006	6.820 ± 0.206
	(.027 ± .006)	(16.324 ± 1.422)

Note. At each sample size, 25 deconvolutions are simulated. The results are evaluated by the integrated squared error (ISE) and the total variation (TV). Values of the mean and standard deviation ($M \pm SD$) of the two error measures are shown × 1,000. The upper bound of the filter's pass band is 15. The expected density is an inverse Gaussian density with M = 300 ms and SD = 100 ms. (The TV value of this density is 6.799.) Entries in parentheses are the mean and standard deviation of the corresponding ISE and TV values of the 25 histogram estimates of the inverse Gaussian density.

We conclude that the FFT deconvolution can successfully extract the decision density estimates from a convolution of normal residual density and an inverse Gaussian density. The results are quite satisfactory when the number of observations reaches a couple of thousand and beyond.

Simulation 2. Smith (1990) found that when a 2-stage gamma distribution was extracted from a 6-stage gamma distribution, the FFT deconvolution method does not yield very satisfactory results. It follows that if one tries to extract an exponential density, the quality of the deconvolved estimate can only be worse. The reason is that the exponential density has a sharp peak; the more sharply peaked a distribution in the time domain, the more frequency components in its discrete Fourier representation must be retained to assure the same level of accuracy. Thus, it makes a good case to demonstrate how the value of the upper bound of the filter affects the results in the FFT deconvolution. In addition, the exponential distribution is the single-stage time process predicted by Ashby and Townsend (1980) and Ashby (1982) in their serial model for reaction time. It is important, theoretically as well as technically, to know how well this distribution can be extracted by deconvolution.

In this simulation, we set the number of trials to be 4,096 and varied the value of the upper bound of the pass band from 7 to 21 with a step size of 2. Smith's (1990) Figure 7 showed that the rectangular filter provided a better fit than the parabolic filter to the abrupt peaking of the 2-stage gamma distribution. To compare results, both the parabolic and rectangular filters are applied to the same set of random observations. Overall, there are 8 (values of filter parameter) × 2 (number of filters) = 16 conditions. In each condition, the deconvolution is iterated 50 times.

We represent the total latency distribution by the convolution of a normal distribution with M = 300 ms and SD = 60 ms and an exponential distribution of M = 100 ms (and SD = 100 ms). For plotting purposes, we add a location shift of 200 ms to the exponential distribution. The shift is equivalent to adding 200 ms to the expected value of the normal component. The deconvolution procedure is applied to the histogram estimates of the time distributions. In parallel, we obtain an average TV value of 22.850 with an SD = 1.310 from 50 histogram estimates of the exponential density. This indicates that a sample of 4,096 trials is sufficiently large for the smoothness requirement of the expected density whose TV value is 20. The ISE values of the

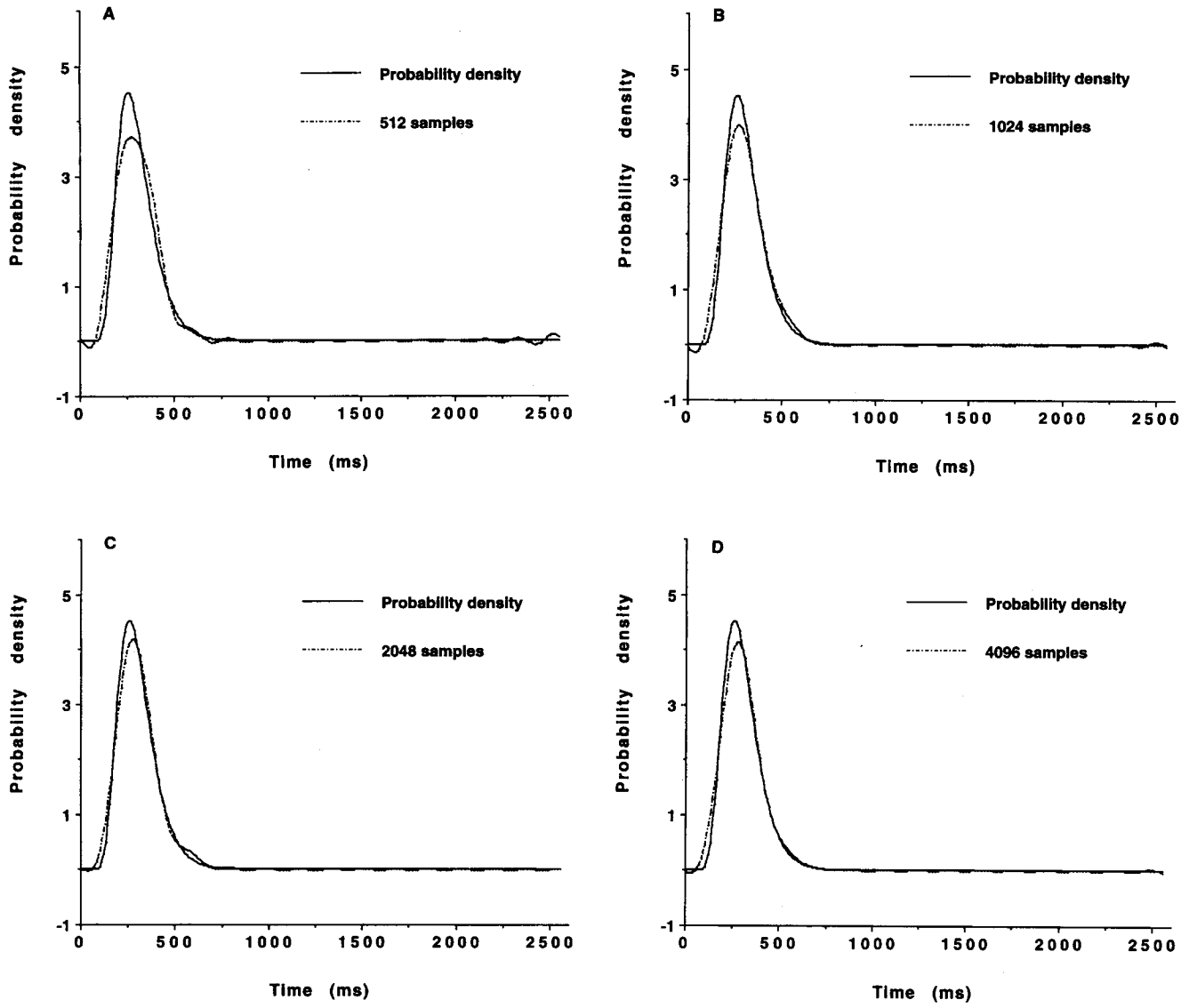

Figure 1. Illustration of the effect of the number of observations on the accuracy of the deconvolved estimates. The sample sizes are 512, 1,024, 2,048, and 4,096 for parts A, B, C, and D, respectively. The sampling interval is 10 ms.

histogram estimates have an average of 1.029 with an SD = .007. The deconvolved results are summarized in Table 2.

Unlike the results of the previous simulation, the deconvolved estimates do not have smaller ISE values relative to their corresponding histogram estimates. In other words, although filtering produces smoother density estimates, they are not more accurate. When the upper bound of the parabolic filter's pass band $\zeta_o = 21$, we extracted estimates that are as smooth as the exponential density according to the TV value, but the results are the least stable in terms of both ISE and TV measures. In the case of the rectangular filter, the best match between the expected TV value and that of the estimates is when the ζ_o is in the range of 13–15. Consistent with Smith's (1990) results, the rectangular filter outperforms the parabolic filter. Figure 3 illustrates typical exponential densities extracted using the two different filters. It is easy to see that both estimates fail to capture the abrupt peaking of the exponential density at the origin (shifted). We examined plots with varying values of ζ_o and found that an improvement of the match near the peak comes at the cost of introducing large oscillations, that is, sacrificing the smoothness of the estimates. Figure 3 reveals that the numerical agreement in TV values between the deconvolved estimates and the true density does not translate to closeness in the shapes of the distributions. The numerical agreement may, in fact, be an artifact of a large oscillatory component compensated by an overly smoothed estimate. This illustrates the importance of always plotting the deconvolved estimates for inspection.

The results of the simulation show that increasing the upper bound of the filter's pass band does not always improve the qual-

ity of the deconvolved estimates. Instead, there is a trade-off between the accuracy at the peak and the overall accuracy of the shape. The smallest ISE value (based on 4,096 trials) in Table 2 is larger by a factor of 20 than the ISE value based on 512 trials in Simulation 1. The result cannot be considered satisfactory. Because the FFT deconvolution performs poorly when the density to be extracted is not smooth, it cannot validate whether the single-stage time process of a serial model is exponentially distributed or not.

Simulation 3. The purpose of this and the next simulation is to examine the problem of mimicking in the deconvolution method. The two simulations differ in what the experimenter believes to be the correct distribution models for the data. The extent of the mimicking in reaction time observations is manufactured by a simple match of the first two central moments between two candidate distributions. This match is motivated by observing that the sample mean and variance are often used to determine the values of parameters of a theoretical distribution in fitting reaction time data. We should point out that two different types of distributions with a close match in mean and variance can have power spectra far apart.

In this simulation, we assume the one-boundary diffusion model and simulate from it a sample of time observations. We deconvolve with a component distribution as if we thought the data came from a two-stage serial model with gamma distributions. It is interesting to see if the deconvolved estimates are interpretable in terms of the serial model.

One might ask, beyond large ISE and TV values, what anomalous features are there in the deconvolved estimates when the FFT deconvolution technique is used inappropriately? Unfortunately, the answer depends, among other things, on how different the distributions are entering Equation 2, the nature of the noise, and the filters and their pass band chosen to carry out the inversion. This is another reason why we resort to a simulation study. Nevertheless, it is possible to give a few qualitative descriptions when the resulting deconvolved estimates are suspect: (a) The presence of pronounced oscillatory components in the tails of the recovered time distribution usually comes with negative spikes. (b) The shape of the extracted density deviates noticeably from the expected one. (c) The mode of the extracted density does not coincide with that of the expected density; the mode may be shifted, too low or too high. (d) Several modes are present when only one is expected. (e) An isolated hump breaks the pattern of small, oscillatory components at the tails. It is, of course, assumed in the prior observations that the expected density is smooth and well behaved. However, these qualitative differences may easily be smoothed away by the filtering operation. If not, they could inflate the ISE and TV values by a substantial amount.[3]

Table 2
Effect of Window Functions and Their Pass Band on the ISE and TV Characteristics of the Deconvolved Exponential Estimates

ζ_0	ISE	TV
	Rectangular filter	
7	1.757 ± 0.004	15.822 ± 0.334
9	1.436 ± 0.006	17.801 ± 0.589
11	1.215 ± 0.011	19.445 ± 1.041
13	1.060 ± 0.020	21.110 ± 2.402
15	0.969 ± 0.049	23.359 ± 4.295
17	1.021 ± 0.175	29.939 ± 9.143
19	1.927 ± 1.308	58.867 ± 30.044
21	12.995 ± 33.067	153.070 ± 173.050
	Parabolic filter	
7	2.314 ± 0.008	7.119 ± 0.063
9	1.917 ± 0.012	8.319 ± 0.093
11	1.632 ± 0.016	9.313 ± 0.127
13	1.419 ± 0.021	10.168 ± 0.214
15	1.256 ± 0.031	10.100 ± 0.346
17	1.131 ± 0.048	11.927 ± 0.497
19	1.048 ± 0.080	13.881 ± 1.490
21	1.089 ± 0.318	21.397 ± 11.328

Note. At each value of ζ_0, 50 simulations are performed to extract an exponential density from a convolution of a normal and an exponential density. Each run is based on 4,096 trials. The mean and standard deviation of integrated squared error (ISE) and total variation (TV) values from the 50 histogram estimates of the exponential densities are 1.029 ± 0.007 and 22.850 ± 1.310, respectively. (The TV value of the density is 20.)

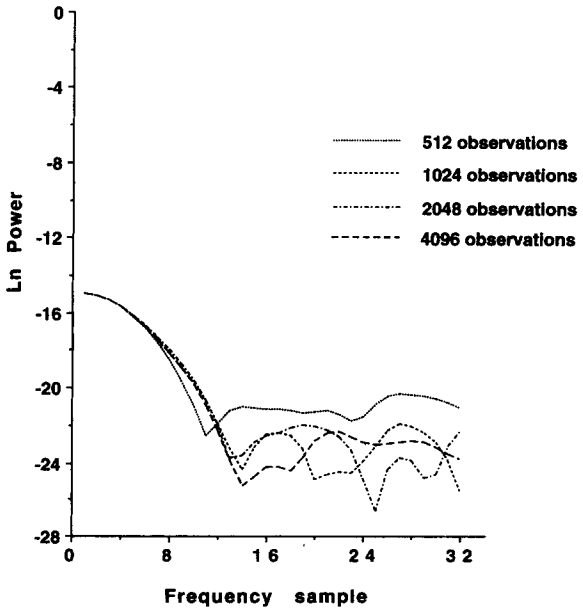

Figure 2. Four sample power spectra of the estimated inverse Gaussian distribution for times. Only the first 32 frequency samples are shown. Note the agreement in the initial portions of the spectra and the divergence beyond the frequency sample 19. Ln = natural logarithm.

[3] Two questions deserve some attention; one theoretical, another practical: (a) Let $\phi_T(\zeta)$ be the characteristic function of an inverse Gaussian distribution and ϕ_R be the characteristic function of a gamma distribution in Equation 1. By inverting the equation in the manner of Equation 2, would the resulting function ϕ_D correspond to any distribution function at all? (b) Suppose there is a real qualitative difference in the deconvolved estimates when no noise is involved. Could the FFT deconvolution technique as it is usually practiced detect such a difference when noise is reintroduced? Because the answer to the first ques-

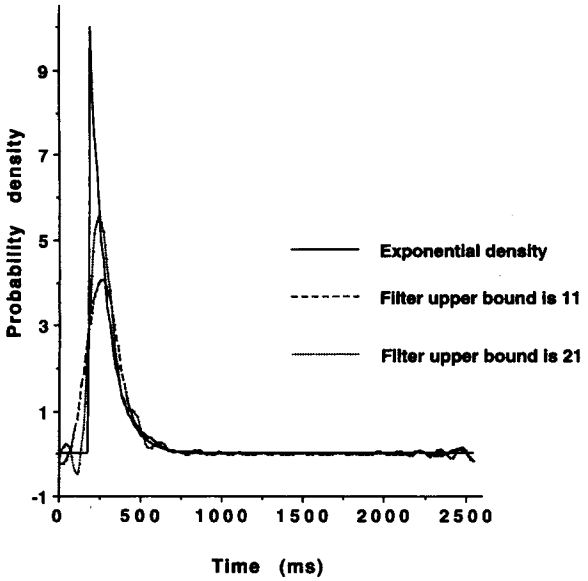

Figure 3. Comparison of deconvolved estimates of an exponential distribution using the rectangular and parabolic filters. The upper bound of the filter's pass band is 13 for the rectangular filter and 21 for the parabolic. The *y*-axis scaling is shown × 1,000.

The decision time predicted by the one-boundary diffusion model follows an inverse Gaussian distribution. We generate a random sample of observations from such a distribution with M = 300 ms and SD = 162 ms. Added to the sampled values (point by point) of the inverse Gaussian variate are values sampled from a normal distribution of M = 300 ms and SD = 60 ms. Thus, the total latency is modeled by a convolution of an inverse Gaussian distribution and a normal distribution (as residual latency). These distribution parameters were chosen because they provide fits of the one-boundary diffusion model to empirical data. A histogram estimate of the total time is shown in Figure 4.

The convolution histogram is well approximated by the simulation from a 4-stage gamma distribution with M = 600 ms and SD = 173 ms. Assuming that the 4-stage gamma density fits the time data well, the experimenter might proceed to interpolate the decision density by deconvolving the total time with a residual density of a 2-stage gamma density. The expected decision density is another 2-stage gamma density. A total of 20 conditions is generated by factorially combining four sample sizes (512, 1,024, 2,048, and 4,096) and five values of the filter's upper bound (7, 11, 15, 19, and 23). At each condition, a gamma density is used to deconvolve the convolution histogram of inverse Gaussian and normal densities. There are 50 simulated deconvolutions in each condition. In parallel, we performed 50 correct (in the experimenter's mind) deconvolutions of a 4-stage gamma density by a 2-stage gamma density. The deconvolution is repeated for each of the three window functions described in Equations 4–6. We used the same histogram estimate of gamma distributions in both correct and incorrect deconvolutions to facilitate comparison of results. Tables 3 and 4 summarize the findings of this simulation.

The numbers in Table 3 are the number of times (out of 50) that the ISE (TV) value of the incorrectly extracted estimate is smaller than the correct estimate. For example, using a Hanning filter with ζ_o = 23 in the deconvolutions based on 512 time observations, in 17 out of 50 simulated trials the incorrectly conceived deconvolution gives a smaller ISE value than the correctly extracted 2-stage gamma density estimate. Similarly, in 19 trials out of 50, the TV value is smaller for the incorrectly conceived deconvolutions. The reason for looking at these numbers is that an experimenter generally favors a smoother deconvolved estimate with a small ISE value over an estimate that has both larger ISE and TV values. Of course, if the experimenters are guided by the belief that the correct estimate is a 2-stage gamma density, then they direct their attention to the true TV value of 8.521 for such a density. Table 4 compares ISE and TV values at selective ζ_o values. These conditions have the largest number of time observations, and the correctly extracted 2-stage gamma estimates are superior to those incorrectly deconvolved estimates. Still, the ISE and TV values between the two are quite close. Among the three filters used, the parabolic filter does best in discriminating the correct estimates from the incorrect ones.

Figure 5 is a typical plot of a correct 2-stage gamma density estimate and an incorrect deconvolved estimate. Both estimates

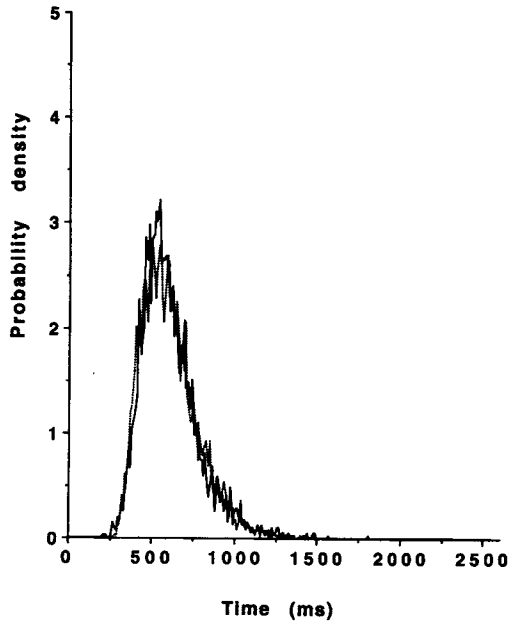

Figure 4. Examples of two hypothetical reaction time distributions. The solid line shows a histogram density estimate of the convolution of an inverse Gaussian density and a normal density. The density of a four-stage gamma random deviate is shown by the dotted line. Each distribution is based on 4,096 simulated observations. The *y*-axis scaling is shown × 1,000.

tion is somewhat technical, we refer the interested readers to the Appendix for a detailed discussion. The answer to the second (in a special case) is examined in Simulations 3 and 4.

Table 3
Results of Deconvolving Inverse Gaussian and Normal Convolutions by Gamma Densities, Part 1

	Filter					
	Hanning		Parabolic		Rectangular	
ζ_0	ISE	TV	ISE	TV	ISE	TV
			$N = 512$			
7	46	3	46	4	8	8
11	40	2	22	12	13	22
15	23	8	14	19	17	20
19	14	18	20	19	22	23
23	17	19	20	20	25	24
			$N = 1,024$			
7	48	1	48	2	6	4
11	47	1	19	5	8	21
15	21	2	7	13	16	27
19	10	8	12	19	22	28
23	10	17	15	23	25	31
			$N = 2,048$			
7	50	0	50	0	2	2
11	47	0	26	4	3	14
15	30	3	2	7	7	26
19	8	3	4	16	18	34
23	3	13	8	30	22	31
			$N = 4,096$			
7	50	0	50	0	0	0
11	50	0	20	0	0	19
15	21	0	0	3	2	29
19	2	0	1	9	9	26
23	0	3	2	17	18	25

Note. The entries are the number of times (out of 50) the integrated squared error (ISE) (total variation [TV]) value of the deconvolved estimate is smaller than the ISE (TV) value of the deconvolved 2-stage gamma density.

miss the peak somewhat and have similar negative loops. It is difficult, however, to make the case that the two estimates shown are from incompatible distributional assumptions. A misinterpretation of the deconvolved results is a real possibility in such a scenario. Though neither component of the true total latency is a combination of exponential stages, the experimenter can still interpret the extracted estimates as rising from a serial model of reaction time with exponentially distributed component processes. Remember that in some uses of the method, the combination and one component are provided, but there is no way of determining whether the component is correct, so the situation in which correct and incorrect decompositions can be discriminated does not arise.

Simulation 4. This simulation complements Simulation 3 by reversing the correct and incorrect distributions. The true total times are observations from a gamma distribution. The experimenter, however, adopts as a model for the reaction time data a convolution of an inverse Gaussian and a normal density. Deconvolution of the total reaction time by a normal component is subsequently carried out, and the resulting estimate is checked against the expected inverse Gaussian density. Will the deviation between the extracted estimate and the expected density be sufficiently large to alert the experimenter that something might be amiss with the present model?

Again, we simulate observations from a correct and incorrect distribution in parallel. A 4-stage gamma distribution with M = 600 ms and SD = 173 ms has the same mean and variance as the sum of a normal distribution with M = 300 ms and SD = 60 ms and an inverse Gaussian distribution with M = 300 ms and SD = 162 ms. We see in Figure 4 that histogram estimates of the two are not noticeably different. Five values (7, 11, 15, 19, and 21) of the upper bound of the filter's pass band are combined with four sample sizes (512, 1,024, 2,048, and 4,096) to create a total of 20 conditions. At each condition, we repeated 50 simulated deconvolutions of a 4-stage gamma density by a normal density. Each deconvolution is implemented once with each of the three filters: Hanning, parabolic, and rectangular. The outcomes are compared with the outcomes from the corresponding correct deconvolutions on the sum of an inverse Gaussian distribution and a normal distribution deconvolved by the normal distribution.

Tables 5 and 6 present a summary of ISE and TV values from the simulation. The numbers in Table 5 show that no matter

Table 4
Results of Deconvolving Inverse Gaussian and Normal Convolutions by Gamma Densities, Part 2

ζ_0	ISE	TV
	Rectangular filter	
7	.282 ± .024	13.895 ± 0.589
	(.207 ± .003)	(11.094 ± 0.623)
11	.292 ± .068	13.415 ± 1.445
	(.096 ± .020)	(12.871 ± 1.890)
15	.378 ± .103	16.359 ± 2.726
	(.135 ± .100)	(17.376 ± 4.383)
	Parabolic filter	
11	.187 ± .020	8.190 ± 0.203
	(.181 ± .015)	(7.492 ± 0.182)
15	.200 ± .042	9.551 ± 0.363
	(.097 ± .014)	(8.601 ± 0.443)
19	.249 ± .056	11.441 ± 0.781
	(.081 ± .028)	(10.411 ± 1.135)
	Hanning filter	
15	.189 ± .021	7.748 ± 0.173
	(.183 ± .015)	(6.917 ± 0.165)
19	.189 ± .033	8.926 ± 0.304
	(.115 ± .014)	(7.797 ± 0.354)
23	.216 ± .044	10.308 ± 0.578
	(.090 ± .018)	(9.801 ± 0.814)

Note. Each entry is the mean and standard deviation ($M \pm SD$) of 50 simulations. Each run is a deconvolution by a 2-stage gamma density of the convolution of an inverse Gaussian and normal density. Each simulation is based on 4,096 trials. The extracted density estimate is compared with a 2-stage gamma density extracted from a 4-stage gamma density shown in parentheses. (The true total variation [TV] value of the 2-stage gamma density is 8.521.) ISE = integrated squared error.

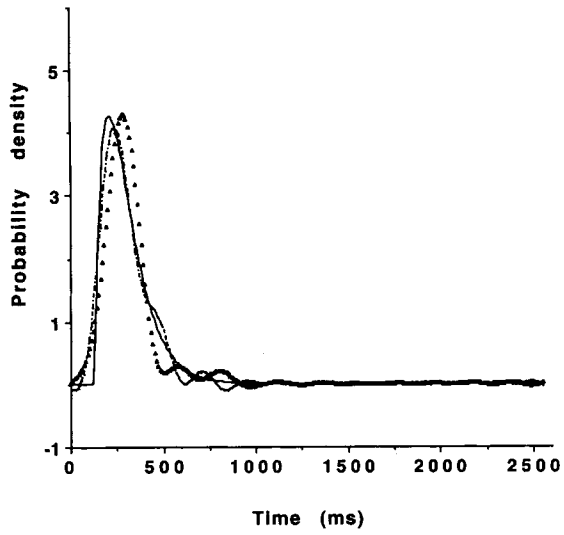

Figure 5. Comparison of the deconvolved estimates of the decision time distribution. The solid line is the expected 2-stage gamma density. The dotted line shows the extracted 2-stage gamma density under a correct deconvolution procedure. The triangular symbols trace out the extracted density estimate under the diffusion model assumption.

which filter is used, the extracted estimates of a gamma density by a normal density are smoother than the correct estimates. Moreover, there is a tendency for the rectangular filter to undersmooth and for the Hanning filter to oversmooth the extracted estimates at the range of cutoff frequencies used. This can be observed by comparing the TV value of the inverse Gaussian density (6.799) with the TV values in Table 6. In general, the correctly extracted estimates have smaller ISE values than the incorrectly extracted estimates. There can be up to a factor of 4 magnitude of difference in ISE value between the two extracted estimates. For example, using a parabolic filter with $\zeta_o = 15$, the extracted inverse Gaussian density estimates seem a lot better than the extracted estimates of a gamma distribution by a normal density based on the same sample size. It is important to find out where the discrepancy lies and whether the two extracted estimates are qualitatively different. Figure 6 shows a typical graph of the two density estimates plotted against time. We added the density function of the inverse Gaussian density for comparison. The shape and the peak location of the inverse Gaussian density are well captured by both estimates. The only obvious gap is the low peak of the incorrect deconvolved estimates.

However, a low peak is not uncommon in deconvolved estimates. We are aware of the true cause of this deviation because we have an exact alternative with which to compare the results. In fact, the ISE values of the incorrect deconvolutions are even slightly better than the results of extracting a 2-stage gamma density from a 4-stage gamma density (see the ISE values in the parenthesis of Table 4). The question is not which deconvolution gives the smallest ISE, but are there ways to do diagnostics on the extracted results? From the results of the simulation, it is not clear what would alert the experimenter to reconsider the assumptions of a one-boundary diffusion model.

Table 5
Results of Deconvolving Gamma Densities by Normal Densities, Part 1

	Filter					
	Hanning		Parabolic		Rectangular	
ζ_0	ISE	TV	ISE	TV	ISE	TV
			$N = 512$			
7	4	44	4	46	9	43
11	4	45	2	46	12	31
15	4	45	2	46	18	20
19	5	34	11	24	22	23
23	10	25	23	24	27	26
			$N = 1{,}024$			
7	1	49	0	49	0	46
11	0	49	1	48	4	32
15	0	49	2	40	20	28
19	2	47	11	29	24	26
23	9	27	20	26	25	27
			$N = 2{,}048$			
7	0	50	0	50	0	50
11	0	50	0	50	0	37
15	0	50	0	47	10	27
19	0	46	4	23	22	28
23	5	27	17	22	19	20
			$N = 4{,}096$			
7	0	50	0	50	0	50
11	0	50	0	50	0	35
15	0	50	0	48	1	16
19	0	48	0	22	19	19
23	1	28	14	20	20	23

Note. The number of times (out of 50) the integrated squared error (ISE) (total variation [TV]) value of the deconvolved estimate is smaller than the ISE (TV) value of the extracted inverse Gaussian density.

One might argue that if experimenters found the preponderance of evidence for Model A (e.g., a serial model) over Model B (e.g., a diffusion model) at different pass bands of various filters and at different sample sizes, then they might be tracking down the correct model. The trouble with this approach is that both models may be wrong. The situation is analogous to using only the coefficient of determination to choose among competing regression models. In Simulations 3 and 4, we are able to create problems for interpreting deconvolved estimates by simply choosing two different distributions and matching their first two central moments approximately. Presumably, there are more sophisticated ways to choose distributions that are close in some other sense.[4]

The findings of Simulations 3 and 4 suggest that the experimenter must obtain independent, corroborative evidence about the shape of the reaction time distributions to avoid misinterpreting the FFT deconvolution results, but at the moment we

[4] For example, one could try to produce a match in the initial portions of the Fourier transforms with two different distributions.

cannot see what form this evidence would take. Perhaps this is a major challenge to any researcher who wishes to decompose a reaction time data set. For the time being, we propose that one should run several deconvolutions on the same data set. Each deconvolution procedure assumes different plausible shapes about the time distributions. The same data set can be bootstrapped to yield sample observations of different sizes. The best distribution can then be selected on the basis of ISE and TV criteria. If the deconvolution procedure is operating under correct assumptions, the ISE value should decrease in an orderly manner as sample size increases. The width of the filter should be varied systematically (within the boundary determined by the power spectrum) until a minimum TV value is reached or the width is close to some value of a theoretical density. It is also important to plot the deconvolved estimates against the competing shape assumptions to detect the locations of the discrepancy between the extracted and expected estimates. This is because the same amount of discrepancy may appear in different places (globally or locally) under different shape assumptions. It is not clear which is worse: missing the peak or having more ripples at the tails.

Conclusion

The FFT deconvolution method is a potentially useful data-processing tool. If the interest of the study is parameter estimation for an *assumed* serial stage model, then our first two simulations and Smith's (1990) example show that the method can work reasonably well, except for the case in which the single-stage distribution, such as the exponential distribution, has a sharp peak. However, the main problem is often not how to fit a model but rather which one of the models can be validated. In this case, our results show that the method alone is not capable of determining the shape of a time distribution in a serial process when the observed time data are statistically consistent with other theories. In fact, the method is independent of the prior model assumptions. The method also does not suggest how one might better analyze a given set of data when the assumed model is wrong. We conclude that the application of the method is limited. Our simulations demonstrate the importance of developing adequate tests on the fundamental assumptions of stage models, that is, the existence of discrete, nonoverlapping, additive, independent serial processes. Without them, the powerful Fourier deconvolution method can give very misleading answers because of the inherent mimicking problem. Thus, the challenge is to obtain independent evidence for the validity of the subtractive method.

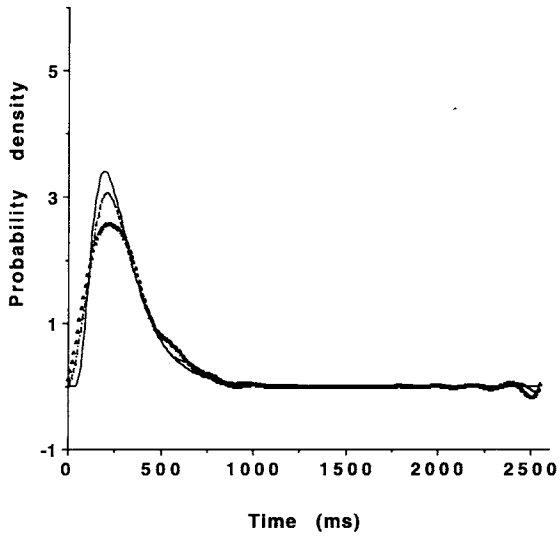

Figure 6. Comparison of the deconvolved estimates of the decision time distribution. The solid line is the expected inverse Gaussian density. The dotted line shows the extracted inverse Gaussian density under an appropriate deconvolution procedure. The triangular symbols illustrate the result of deconvolving a 4-stage gamma distribution by a normal distribution.

Table 6
Results of Deconvolving Gamma Densities by Normal Densities, Part 2

%	ISE	TV
	Rectangular filter	
7	.093 ± .008	7.577 ± 0.293
	(.068 ± .001)	(8.664 ± 0.271)
11	.082 ± .020	8.258 ± 0.068
	(.021 ± .006)	(8.945 ± 0.929)
15	.173 ± .067	13.663 ± 3.623
	(.046 ± .031)	(11.757 ± 2.468)
	Parabolic filter	
11	.115 ± .011	5.647 ± 0.100
	(.060 ± .007)	(6.144 ± 0.112)
15	.094 ± .015	6.162 ± 0.259
	(.026 ± .007)	(6.695 ± 0.186)
19	.129 ± .033	9.223 ± 2.027
	(.031 ± .016)	(8.901 ± 1.464)
	Hanning filter	
15	.125 ± .011	5.258 ± 0.105
	(.065 ± .007)	(5.786 ± 0.110)
19	.106 ± .014	5.705 ± 0.213
	(.036 ± .008)	(6.271 ± 0.197)
23	.168 ± .277	10.297 ± 12.953
	(.109 ± .478)	(11.098 ± 16.821)

Note. Each entry is the mean and standard deviation of 50 simulated deconvolutions of a 4-stage gamma density by a normal density (based on 4,096 trials). The extracted density is compared with the expected inverse Gaussian density having a total variation (TV) value of 6.799. The entries in the parentheses are the corresponding mean and standard deviation of integrated squared error (ISE) and TV values of the inverse Gaussian extracted from a convolution of a normal and an inverse Gaussian density.

References

Ashby, F. G. (1982). Testing the assumptions of exponential, additive reaction time models. *Memory and Cognition, 10,* 125–134.

Ashby, F. G., & Townsend, J. T. (1980). Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology, 21,* 93–123.

Bendat, J. S., & Piersol, A. G. (1971). *Random data: Analysis and measurement procedures.* New York: Wiley.

Bloxom, B. (1979). Estimating an unobserved component of a serial model. *Psychometrika, 44,* 473–484.

Brigham, E. O. (1988). *The fast Fourier transform and its applications.* Englewood Cliffs, NJ: Prentice-Hall.

Burbeck, S. L., & Luce, R. D. (1982). Evidence from auditory simple reaction times for both change and level detectors. *Perception and Psychophysics, 32,* 117–133.

Chhikara, R. S., & Folks, J. L. (1989). *The inverse Gaussian distribution: Theory, methodology, and applications.* New York: Marcel Dekker and Basel.

Donders, F. C. (1969). On the speed of mental processes. In W. G. Koster (Ed. and Trans.), *Attention and performance* (Vol. 2, pp. 412–431). Amsterdam: North-Holland. (Original work published in 1868–1869).

Dzhafarov, E. N. (1992). The structure of simple reaction time to step-function signals. *Journal of Mathematical Psychology, 36,* 235–268.

Dzhafarov, E. N., & Schweickert, R. (1994, August). *Simple reaction time decomposition: A general theory (almost).* Paper presented at the 27th annual Mathematical Psychology Meeting, University of Washington, Seattle.

Feller, W. (1966). *An introduction to probability theory and its applications* (Vol. 2). New York: Wiley.

Green, D. M. (1971). Fourier analysis of reaction time data. *Behavior Research Methods and Instrumentation, 3,* 121–125.

Green, D. M., & Luce, R. D. (1971). Detection of auditory signal presented at random times: III. *Perception and Psychophysics, 9,* 257–268.

Hamming, R. W. (1983). *Digital filters* (2nd ed.). Signal processing series. Englewood Cliffs, NJ: Prentice-Hall.

Kohfeld, D. L., Santee, J. L., & Wallace, N. D. (1981). Loudness and reaction time. II. Identification of detection components at different intensities and frequencies. *Perception and Psychophysics, 29,* 535–549.

Luce, R. D. (1986). *Response time.* Oxford, England: Oxford University Press.

McCullagh, P. (1994). Does moment generating function characterize a distribution? *American Statistician, 48,* 208.

Michael, J. R., Schucany, W. R., & Haas, R. W. (1976). Generating random variables using transformation with multiple roots. *American Statistician, 30,* 88–90.

Oberhettinger, F. (1973). *Fourier transform of distributions and their inverses. A collection of tables.* New York: Academic Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1992). *Numerical recipes. The art of scientific computing* (2nd ed.). London: Cambridge University Press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R. (1988). A note on mimicking additive reaction time models. *Journal of Mathematical Psychology, 32,* 192–204.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance* (Vol. 14, pp. 611–653). Cambridge, MA: MIT Press.

Rouder, J. (1993, August). *Towards a nonasymptotic, distribution-free decomposition of simple reaction time.* Paper presented at the 26th annual Mathematical Psychology Meeting, University of Oklahoma, Norman.

Shaffer, J. P., Shaughnessy, E. J., & Jones, P. L. (1984). The deconvolution of Doppler-broadened annihilation measurements using fast Fourier transforms and power spectral analysis. *Nuclear Instruments and Methods in Physics Research, 5,* 75–79.

Silverman, H. F., & Pearson, A. E. (1973). On deconvolution using the discrete Fourier transform. *IEEE Transactions on Audio and Electroacoustics, 2,* 112–118.

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

Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology, 25,* 168–197.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1,* 46–54.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes.* Cambridge, England: Cambridge University Press.

Turchin, V. F., Kozlov, V. P., & Malkevickz, M. S. (1971). The use of mathematical-statistics methods in the solution of incorrectly posed problems. *Uspekhi Fizicheskih Nauk, 13,* 681–703.

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time distributions: Mixtures and parameter variability. *Psychonomic Bulletin and Review, 2,* 20–54.

(Appendix continues on next page)

Appendix

Responses to Questions Posed in Footnotes

The theoretical question we posed in Footnote 3 is whether the ratio of the characteristic function of an inverse Gaussian distribution and the characteristic function of a gamma distribution is or is not a characteristic function. If it is not, then deconvolution in the manner prescribed by Equation 2 results in anything except a proper density function. This is because a distribution function is uniquely determined by its characteristic function and vice versa (see, e.g., Feller, 1966). From Equations 10 and 12, we form a function with argument $-\infty < \zeta < \infty$ and four parameters: $\mu > 0$, $\lambda > 0$, $\alpha > 0$, and r is a positive integer,

$$\phi(\zeta) = \frac{\exp\left((2\lambda)^{1/2}\left[\left(\frac{\lambda}{2\mu^2}\right)^{1/2} - \left(\frac{\lambda}{2\mu^2} - i\zeta\right)^{1/2}\right]\right)}{\left[1 - \left(\frac{i\zeta}{\alpha}\right)\right]^{-r}}.$$

It is, in general, difficult to demonstrate if an arbitrary function is or is not a characteristic function. Our strategy is to choose a particular set of values for the parameters and check if the modulus of the function $\phi(\zeta)$ is less than or equal to one for all values of ζ. This is a condition any characteristic function must satisfy (see, e.g., Feller, 1966). We set $\mu = \lambda = 2$, $\alpha = 1$, and $r = 2$. (This corresponds to an inverse Gaussian distribution with both $M = 2$, $SD = 2$, and a 2-stage gamma distribution with $M = 2$ and $SD = \sqrt{2}$.) Thus, the modulus of $\phi(\zeta)$ simplifies to

$$|\exp[1 - 2(1/4 - i\zeta)^{1/2}](1 - i\zeta)^2|.$$

We compute the modulus of the product separately, because the modulus of a product is equal to the product of the modulus of the factors. A trite calculation shows that the modulus of the first factor reduces to

$$\exp\left(1 - 2\sqrt{\frac{1/4 + \sqrt{1/16 + \zeta^2}}{2}}\right)$$

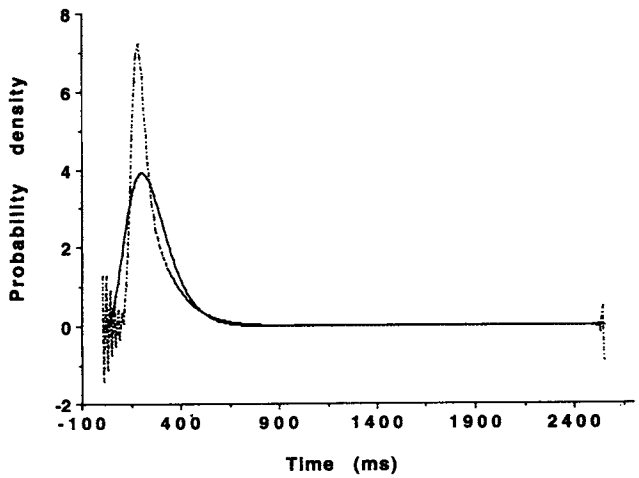

Figure A1. For the two deconvolved estimates, the solid line represents the estimate of a 5-stage gamma distribution from deconvolving a 7-stage gamma distribution from a 12-stage gamma distribution, and the dashed line represents the estimate of a function obtained from deconvolving a 7-stage gamma distribution from an inverse Gaussian distribution.

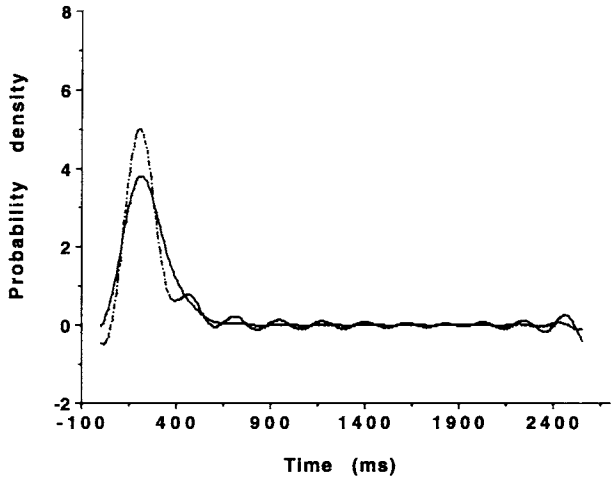

Figure A2. The same two deconvolved estimates from Figure A1, but a filter is used before inversion.

whereas the modulus of the second factor becomes $(1 + \zeta^2)$. Therefore, $1 < |\phi(1)| \approx 1.0971$. We conclude (only) that the function obtained from deconvolving an inverse Gaussian distribution with $M = 2$ and $SD = 2$ by a 2-stage gamma distribution with $M = 2$ and $SD = \sqrt{2}$ is not a proper density function. (With the gamma distribution mean smaller than the inverse Gaussian mean, the tails of the ratio of the characteristic function rise after initially decreasing from 1 in a W shape.) This example also shows that the ratio of the characteristic function of an inverse Gaussian distribution and the characteristic function of a gamma distribution is not automatically a characteristic function.

In response to the practical question we posed in Footnote 3, we provide a graphic illustration of how filtering operation in the FFT deconvolution procedure can remove the possibility of detecting a difference between the correct and incorrect deconvolved estimates.

Figure A1 shows two deconvolved estimates. The solid line is the estimate of a 5-stage gamma distribution obtained from deconvolving a 7-stage gamma distribution from a 12-stage gamma distribution ($\alpha = 0.02$). The dashed line is the estimate of a function obtained from deconvolving a 7-stage gamma distribution from an inverse Gaussian distribution whose first two central moments coincide with that of a 12-stage gamma distribution with $\alpha = 0.02$. The 256 histogram estimates, each 10 ms apart, are calculated from the formulas of the corresponding probability density functions. No filter is used. Notice the salient oscillatory components with negative spikes at the origin of the recovered "incorrect" estimate, a less noticeable one at the right tail, the sharply peaked mode, and the deviation of the shape from the expected 5-stage gamma density. (The vertical axis scaling is shown × 1,000.)

Figure A2 shows the results of the same deconvolution procedures as in Figure A1, except that a parabolic filter is used before inversion. The upper bound of the filter's pass band is 23. Compared with Figure A1, we observe that as the oscillations at the origin of the incorrect estimate are smoothed away and are replaced by a negative region, the shape becomes more similar to the expected 5-stage gamma density. Though filtering introduces oscillatory components to both correct and incorrect estimates, those on the right tail of the incorrect one have a much larger magnitude. However, these loops are not unusual (see Smith's, 1990, Figure 7 for comparison).

Figure A3 shows the results of the same deconvolution procedures as in Figure A2. The same filter is used. However, the 12-stage gamma distribution and inverse Gaussian distribution are simulated. Each is based on 4,096 random observations. Histogram estimates were obtained by grouping the observations into 256 bins, each 10 ms wide, and normalizing. Filtering is performed using the same parabolic window function (with the same pass band) as before. The noise in random trials introduces even more pronounced negative loops to the two estimates. Encountering such a deconvolved estimate, the experimenter's natural reaction is either to try another filter, to reduce the upper bound of the filter's pass band, or both.

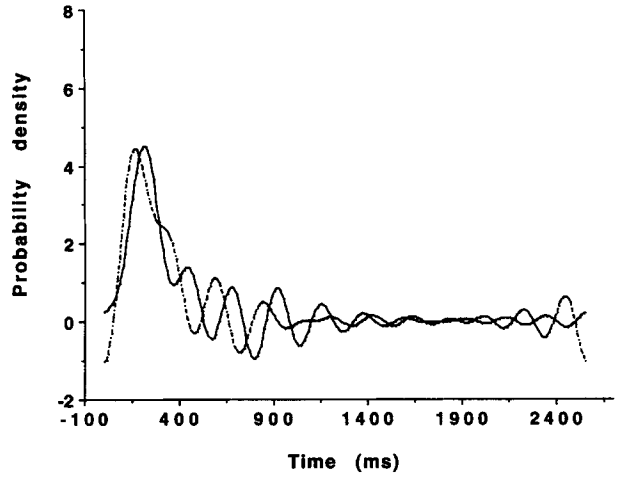

Figure A3. The results of the same deconvolution procedures as in Figure A2 and the same filter are used, but the 12-stage gamma distribution and inverse Gaussian distribution are simulated.

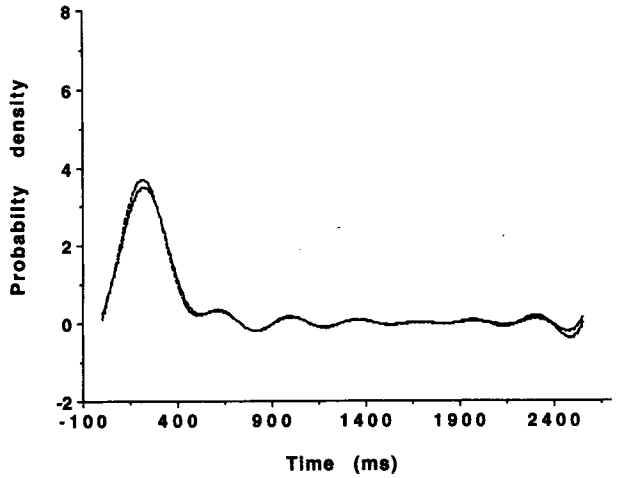

Figure A4. The results of the same deconvolution estimates as in Figure A3, but there is no longer any qualitative difference between the two deconvolution estimates.

Figure A4 shows the results of the same deconvolved estimates as in Figure A3. The upper bound of the filter's pass band is reduced from 23 to 15. Compared with Figure A1, there is no longer any qualitative difference between the two deconvolved estimates. In summary, with filtering in the frequency domain, the oscillations in the time domain are dampened and one cannot distinguish from which distribution, inverse Gaussian or gamma, the reaction time data is simulated.

Received September 7, 1993
Revision received December 13, 1994
Accepted December 13, 1994 ■