Annals of Biomedical Engineering, Vol. 18, pp. 629–654, 1990
Printed in the USA.

0090-6964/90 $3.00 + .00

The Identification of Nonlinear Biological Systems: Wiener Kernel Approaches

Michael J. Korenberg

Department of Electrical Engineering
Queen's University
Kingston, Ontario, Canada

Ian W. Hunter

Department of Biomedical Engineering
and McGill Research Centre for Intelligent Machines
McGill University
and the Canadian Institute for Advanced Research
Montreal, Canada

(Received 12/6/88; Revised 6/13/89)

Detection, representation, and identification of nonlinearities in biological systems are considered. We begin by briefly but critically examining a well-known test of system nonlinearity, and point out that this test cannot be used to prove that a system is linear. We then concentrate on the representation of nonlinear systems by Wiener's orthogonal functional series, discussing its advantages, limitations, and biological applications. System identification through estimating the kernels in the functional series is considered in detail. An efficient time-domain method of correcting for coloring in inputs is examined and shown to result in significantly improved kernel estimates in a biologically realistic system.

Keywords *— System identification, Wiener kernels, Nonlinear systems.*

INTRODUCTION

It is almost trite to state that all biological systems exhibit some form of nonlinear behavior, such as saturation, in response to certain inputs. However, it is not a trivial problem to decide what form of nonlinear model might be appropriate to describe the system's response to a restricted class of inputs. Indeed if the system or its

---

Acknowledgment — This research was supported by grants from the Medical Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, and the Advisory Research Committee of Queen's University. IWH would like to thank the Canadian Institute for Advanced Research for personal support. We thank Dr. V.Z. Marmarelis for pointing out the obstacle to calculating Wiener kernels for a Hammerstein model, and the referees for their astute comments.

Address correspondence to M.J. Korenberg, Department of Electrical Engineering, Queen's University, Kingston, Ontario, Canada K7L 3N6.

response is noisy, it is often difficult even to decide whether or not the system is linear.

The techniques presently available for the identification of nonlinear systems can basically be grouped into three classes:

1. kernel, functional series, or nonparametric approaches (e.g., Wiener and Volterra representations);
2. cascade and block structured approaches (e.g., Hammerstein, Wiener, Uryson, LNL, and parallel cascade structures);
3. parametric approaches (e.g., bilinear, state-affine, and nonlinear difference equation models).

Recently we have considered cascade approaches to the identification of nonlinear systems (26,40). Parallel cascade models for discrete-time systems were examined by Palm (53) and identification using such a model was undertaken by Korenberg (30). Billings (4) and coworkers [e.g., (45)] have summarized some of their extensive work on system identification via nonlinear difference equation models. Orthogonal identification of nonlinear difference equation models for systems of unknown structure has been performed recently (31,35,36,51).

In this paper we examine some of the techniques available for the identification of nonlinear systems via the Wiener kernel approach (see 46,56,69,70). Since nonlinear systems are in general more arduous to identify than their linear counterparts, several authors (e.g., 6,7,20,75) have proposed simple preliminary investigations to determine whether the system under study is nonlinear. Thus we start by considering a well known test for system nonlinearity. We then proceed to the Wiener kernel methods after a brief introduction to the Volterra functional series.

TESTS FOR SYSTEM NONLINEARITY

Billings and Voon (6,7) have recently proposed tests for system nonlinearity that are attractively simple and apparently very general. While these tests are very useful, they have sometimes been misinterpreted as providing sufficient conditions for linearity whereas in fact they generally constitute only necessary conditions. For example consider the following test.

Let $u(t)$ be a signal with zero mean (first-order) and zero third-order moments and suppose that all even-order moments exist. Billings and Voon (7) give as examples of such signals a sine wave, and a Gaussian or ternary sequence. Let

$$x(t) = u(t) + b \ , \tag{1}$$

where $b \neq 0$, be the input to the system to be tested and let $y(t)$ be the corresponding output. Define

$$\phi_{y'y'^2}(\tau) \equiv \overline{(y(t-\tau) - \bar{y})(y(t) - \bar{y})^2} \ . \tag{2a}$$

Here the primes are used to denote that the mean has been subtracted from the signal and overbars denote (infinite) time averages. Then Billings and Voon (6,7) assert that if the system is linear then the higher-order auto-correlation function

$$\phi_{y'y'^2}(\tau) = 0 \quad \forall \tau \ . \tag{2b}$$

Note that the test is a function only of the output. This test has become popular and for example Varlaki *et al.* (75) have praised it as being particularly simple and straightforward to apply. Indeed it is easy to show that every linear system must satisfy Eq. 2b and to this extent the test is extremely useful and constitutes a valuable contribution by Billings and Voon. Thus, if Eq. 2b is not satisfied (in practice, within specified confidence limits, since only finite-length time averages can actually be used) then there is a corresponding probability that the system is nonlinear, and the wider the confidence limits the greater the probability. However the test has been misused in that the converse (namely, if Eq. 2b holds for all τ then the system must be linear) has sometimes been asserted, which is incorrect. Thus, contrary to such an assertion, it is possible for a nonlinear system to satisfy Eq. 2b for all τ. In short, the fact that a system satisfies Eq. 2b says nothing about whether it is linear or nonlinear, as we show below. Before considering this we emphasize that b in the Billings and Voon test (see Eq. 1) is nonzero. Otherwise it would be trivial to construct a counter-example to the notion that satisfying Eq. 2b is sufficient for linearity. For example, let the input be $x(t) = \sin \omega t$, and the nonlinear system be a simple cuber. Then the output, $y(t)$, would simply be the sum of sinusoids of frequencies ω and 3ω (odd-order harmonics and therefore of nonzero frequency) so that the output mean, $\bar{y}$, would equal zero. Hence, the right side of Eq. 2a would be the infinite time average of sinusoids of odd-order harmonics and thus equal zero. However the Billings and Voon test does not permit use of a simple sinusoid without an offset, since $b \neq 0$. Hence, the right side of Eq. 2a will include the time average of even-order harmonics (including the zeroth-order harmonic) and thus it is not clear that this time average would equal zero for all τ. Therefore a more carefully constructed counter-example (to the belief that Eq. 2b suffices for linearity) is warranted, and this is presented after a final brief comment.

It may seem obvious to the reader that one could not prove that a system is linear using only a single test input. Might not a highly nonlinear system exist whose response to the particular input is zero? This is indeed a reasonable question, especially if no restrictions were placed on the allowable class of nonlinear systems. However, in this paper we confine ourselves to systems having a Volterra or a Wiener functional expansion (see below). Moreover, we assume that a sufficiently "rich" test input is used, for example that u in Eq. 1 is white Gaussian noise. The response to such a test input fully characterizes a nonlinear system. If the resulting system output were zero, then *all* the Wiener kernels (discussed below) would be zero, and the system response to any other input would be zero "almost everywhere."

Turning to our counter-example, we now construct a simple nonlinear system (having both a Volterra and a Wiener functional expansion) which satisfies Eq. 2b even when the test input is sufficiently rich to identify the system.

We will assume that $u(t)$ in Eq. 1 is zero-mean white Gaussian noise. This clearly satisfies Billings and Voon's requirement of having zero mean and zero third-order moments, as these authors have noted. Suppose that the system to be tested has the

Wiener model structural form, that is, a dynamic linear system with impulse response, $h(\tau)$, followed by a static nonlinearity. Furthermore suppose that

$$\int_0^\infty h(\tau)\,d\tau = 0 \ , \tag{3}$$

and that the static nonlinearity is an odd function. A linear system satisfying Eq. 3 is often termed a high-pass system (having a zero steady-state response to a step input). For example

$$h(\tau) = e^{-\tau} - 2e^{-2\tau}, \quad \tau \geq 0 \ ,$$

clearly satisfies Eq. 3. As shown below, the use in our counter-example of an impulse response satisfying Eq. 3 causes the nonzero offset b (in Eq. 1) to have no effect. For simplicity we will assume that the odd function for the static nonlinearity is a simple cuber so that the system output is

$$y(t) = \left[\int_0^\infty h(\tau)x(t-\tau)\,d\tau\right]^3 \ . \tag{4}$$

It follows readily from Eq. 1 that

$$\int_0^\infty h(\tau)x(t-\tau)\,d\tau = \int_0^\infty h(\tau)u(t-\tau)\,d\tau \ , \tag{5}$$

in view of Eq. 3. Consequently x in Eq. 4 can be replaced by u so that the system output is

$$y(t) = \left[\int_0^\infty h(\tau)u(t-\tau)\,d\tau\right]^3 \ . \tag{6}$$

Therefore $\bar{y} = 0$, and the right side of Eq. 2a reduces to

$$\phi_{y'y'^2}(\tau) = \overline{y(t-\tau)y^2(t)} \ . \tag{7}$$

Recall that our choice for $u(t)$ was zero-mean Gaussian noise. It follows from Eq. 6 that the right side of Eq. 7 is the infinite time average of an odd number (nine) of u factors and therefore equals 0 $\forall\tau$. [Note that this result depends on the stimulus having a symmetric probability density function (zero odd-order moments).] Thus our nonlinear Wiener model satisfies Eq. 2b and is a counter-example to the notion that this equation is a sufficient condition for linearity when the test input is as defined above.

Our Wiener model also provides a counter-example to the belief that another proposed test from the literature suffices to establish linearity. The test (which can be used to establish nonlinearity) requires that the signal u in Eq. 1 belong to the separable class of random processes. Let $f(u_1, u_2; \tau)$ be the second-order probability density function of the process u(t). Define

$$g(u_2,\sigma) = \int_{-\infty}^{\infty} u_1 f(u_1,u_2;\sigma)\, du_1 \ .$$

If there exist functions g_1, g_2 such that $g(u_2,\sigma) = g_1(u_2)g_2(\sigma)$ for all u_2 and σ, then $u(t)$ is called a separable process (4,5). The Gaussian process, sine-wave process, and pseudo-random binary sequence are all separable processes (4). If the input u to the system to be tested is a separable process, then it has previously been asserted that the system is linear if and only if

$$\phi_{u^2y'}(\tau) \equiv \overline{u^2(t-\tau)(y(t)-\bar{y})} = 0 \quad \forall\tau \ .$$

Again, for a counter-example consider any Wiener model whose linear system satisfies Eq. 3 and whose static nonlinearity is an odd function. Such a nonlinear system will erroneously appear to be linear according to the immediately above test, when u in Eq. 1 is a zero-mean Gaussian input.

The fact that Eq. 2b is not a sufficient condition for linearity is illustrated in the following simulation.

TESTING FOR LINEARITY: A PITFALL

Our simulation example used a simple Wiener model system consisting of a high-pass, dynamic linear subsystem satisfying Eq. 3 followed by a simple cuber static nonlinearity.

The input was a 1,000 sample unity-variance, zero-mean Gaussian white-noise signal (generated using the technique presented by Hunter and Kearney (23)) to which was added a unity offset. The simulation in this example was carried out using the NEXUS language for simulation, system and signal analysis (25).

A one-dimensional correlation function $\phi_{x'y'}$ can be normalized (7) according to the formula

$$\hat{\phi}_{x'y'}(k) = \frac{\frac{1}{N}\sum_{n=1}^{N-k} (x(n)-\bar{x})(y(n+k)-\bar{y})}{(\phi_{x'x'}(0)\phi_{y'y'}(0))^{.5}} \ .$$

This procedure was used to normalize the correlation function defined by Eq. 2a. The result is plotted in Fig. 1 and lies within the 95% confidence limits of $\pm\ 1.96/\sqrt{N}$ used for the test. Thus our Wiener model would erroneously appear to be linear, so that this test does not provide a sufficient condition for linearity.

VOLTERRA FUNCTIONAL SERIES

Consider a nonlinear, continuous, finite-memory, time-invariant, single-input, single-output, physically realizable (i.e., causal) system. According to the theorem by Frechet (12), such a system can be uniformly approximated, over a uniformly bounded equicontinuous set of inputs, to an arbitrary degree of accuracy by a Volterra (1,2, 80,81) series of sufficient but finite order. If the system is single-input single-output and the finite order is I then the series takes the form:

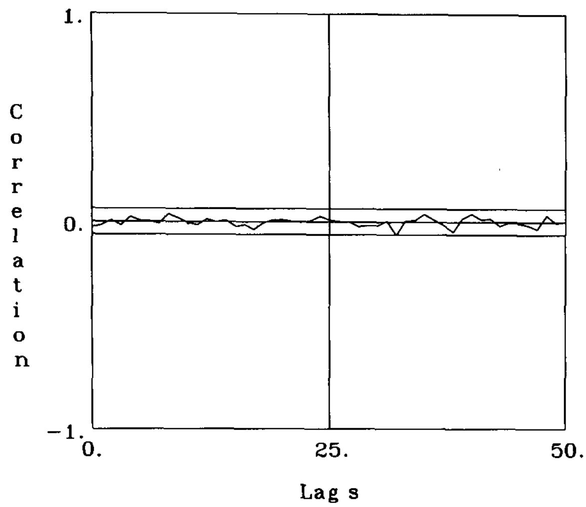

FIGURE 1. Plot of the (normalized) higher-order auto-correlation function (of Eq. 2a) (sometimes used in the past for testing whether a system is linear). The plot lies totally within the indicated 95% confidence limits, suggesting that the simulated system is linear, whereas it is in fact a nonlinear Wiener model. In this and remaining figures the lag is given in seconds, s.

$$\sum_{i=0}^{I} \int_0^{\infty} \cdots \int h_i(\tau_1, \ldots, \tau_i) x(t - \tau_1) \ldots x(t - \tau_i)\, d\tau_1 \ldots d\tau_i$$

$$= h_0 + \int_0^{\infty} h_1(\tau_1) x(t - \tau_1)\, d\tau_1 + \iint_0^{\infty} h_2(\tau_1, \tau_2) x(t - \tau_1) x(t - \tau_2)\, d\tau_1\, d\tau_2$$

$$+ \ldots + \int_0^{\infty} \cdots \int h_I(\tau_1, \ldots, \tau_I) x(t - \tau_1) \ldots x(t - \tau_I)\, d\tau_1 \ldots d\tau_I\,. \quad (8)$$

Here $x(t)$ is the system input and the h_i are the Volterra kernels, essentially multidimensional weighting functions in the Volterra series. In practice the upper limits of the integrals are set equal to T, the duration of the finite memory of the system. Identification of a nonlinear system by a functional expansion essentially reduces to estimation of the Volterra kernels in Eq. 8. Suppose that least-square estimates of the kernels are desired (i.e., estimates minimizing the mean-square error between the system output, $y(t)$, and the predicted output, given by the series of Eq. 8). Then, directly obtaining these estimates requires solution of a simultaneous set of integral equations. The simultaneous set of equations can be avoided by using an orthogonal framework, following Wiener's (83) approach.

WIENER KERNEL APPROACH

In the late 1940s Norbert Wiener (reported by Lee (43)) realized the practical limitations of the nonorthogonal representation. In the most general form, the Volterra series is an infinite series (when I in Eq. 8 is infinite). By using the Gram-Schmidt orthogonalization technique [e.g., (73)] Wiener orthogonalized the Volterra series, under the important assumption that the input is a Brownian process, which is the integral of a white (i.e., has a flat power spectrum) Gaussian process. It is usual now to present Wiener's derivation in terms of a white Gaussian input. The resulting series, known as the Wiener series, can be written

$$y(t) = \sum_{i=0}^{\infty} G_i[k_i,x] \ , \tag{9}$$

where the G_i are orthogonal. That is, if the input, x, is Gaussian white noise with a mean of zero and a given power spectral density, then

$$\begin{aligned} \overline{G_i[k_i,x]G_j[k_j,x]} \quad &= 0 \text{ if } i \neq j \\ &> 0 \text{ if } i = j \ . \end{aligned}$$

G_i is the ith order Wiener G-functional, where the G denotes that the functionals have been orthogonalized with respect to a particular stationary, Gaussian, white input process.

Suppose that the power spectral density of the particular Gaussian white input is P so that $\overline{x(t)x(t-\epsilon)} = P\delta(\epsilon)$ where δ is the unit impulse function of Dirac. Then the first few Wiener functionals are given via Gram–Schmidt orthogonalization as:

$$G_0[k_0,x] = k_0 \tag{10}$$

$$G_1[k_1,x] = \int_0^{\infty} k_1(\tau_1)x(t-\tau_1)\,d\tau_1 \tag{11}$$

$$G_2[k_2,x] = \iint_0^{\infty} k_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)\,d\tau_1 d\tau_2 - P\int_0^{\infty} k_2(\tau,\tau)\,d\tau \tag{12}$$

$$\begin{aligned} G_3[k_3,x] = &\iiint_0^{\infty} k_3(\tau_1,\tau_2,\tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)\,d\tau_1 d\tau_2 d\tau_3 \\ &- 3P\iint_0^{\infty} k_3(\tau_1,\tau_2,\tau_2)x(t-\tau_1)\,d\tau_1 d\tau_2 \ , \end{aligned} \tag{13}$$

where k_i is termed the ith order Wiener kernel.

Note:

1. The Wiener kernels are not in general the same as the Volterra kernels of corresponding order. For example the zeroth-order Volterra kernel is simply the system output when the system input is zero, whereas the zeroth-order Wiener kernel is the mean output for the particular white Gaussian input used. However the corresponding first- and second-order Volterra and Wiener kernels are identical when the system has no higher-order kernels. Suppose that the nonlinear system is a cascade of a dynamic linear, a static nonlinear, and a second dynamic linear component (i.e., an LNL system; see [29,40]). Then the Wiener kernels are directly proportional to the Volterra kernels of corresponding or-

der, assuming both Wiener and Volterra series representations exist. (The Volterra series representation will exist, for example, if the static nonlinear component in the LNL cascade is describable by a polynomial or a power series.)

2. It is frequently claimed that the Wiener series is more general than the Volterra series because an infinite or finite Volterra series exists only for an analytic system (i.e., in brief, a system for which certain functional derivatives of all orders are defined). The kind of convergence displayed by a Volterra series representation is pointwise in stimulus space. A Wiener series is an orthogonal series and so converges in the mean in the sense that the mean square error can be made arbitrarily small by including a sufficient number of Wiener functionals in the truncated series approximation of the output. (That is, the kind of convergence displayed by a Wiener series is L^2.) This explains why it has been claimed that a greater breadth of functionals is treated by the Wiener approach, at the expense of weaker convergence. For example, systems with deadzones (e.g., saturation nonlinearities that have regions where changes in the input do not result in changes in the output) can still have an infinite Wiener series representation while not being exactly describable by a corresponding Volterra series. However this important theoretical distinction has less practical significance because identifying a physical nonlinear system via the functional approach always involves an approximation by a finite number of terms. All that can be obtained in practice is a truncated Wiener series which can always be transformed into a truncated Volterra series providing the same goodness of fit.
3. The Hammerstein model (comprising a static nonlinearity followed by a dynamic linear system) cannot in general be represented by the Wiener functional expansion. As an example consider a squarer followed by a dynamic linear system. If the input to this Hammerstein system is Gaussian white noise (as is required for use of the Wiener expansion) then the output of the squarer will have an infinite mean. This results in an infinite zeroth-order Wiener G-functional for our Hammerstein system and an infinite term in the proposed second-order Wiener G-functional (the first-order kernel is zero in this example). On the other hand Hammerstein systems in which the static nonlinearity is describable by a polynomial do have Volterra series representations. This shows that the class of systems describable by the Wiener expansion is not, at least in one sense, wider than the class describable by a Volterra series (as is frequently contended) since the Wiener series class does not include the Volterra series class as a subset. Indeed there are an infinite number of systems (e.g., many Hammerstein models) having Volterra expansions which do not have Wiener expansions. Such systems may however have a generalized Wiener expansion, known as the Fourier–Hermite expansion, developed for a colored Gaussian input (see 5 below). Moreover, the theoretical difficulty of calculating the Wiener series for a Hammerstein model disappears when time is discretized, and so is not a practical issue.
4. As is well known, the Gaussian input is unbounded (because the Gaussian probability density function is nonzero from minus to plus infinity), and the ordinary derivative of a white Gaussian signal is not defined. Indeed, theoretically the input used to identify the Wiener kernels is not from the set of uniformly bounded equicontinuous signals referred to previously in connection

with Frechet's theorem on approximation by finite order Volterra series. Again this is a distinction without much practical relevance because any physical input is bounded and sufficiently smooth, and lasts only over a finite interval of time. (In practice, of course, system identification generally uses sampled versions of the input and output, forming discrete time signals, and for such signals the truncated functional series becomes a multidimensional polynomial. Most importantly, there are relaxed conditions for convergence of successive polynomial approximations to the system.)

5. Yasui (84,85) has considered generalizations of the Wiener series for colored Gaussian inputs, using "Fourier-Hermite" kernels. This work is related to Schetzen's (68) examination of identification using colored Gaussian signals, and to the work of Victor and Knight (78).
6. The Wiener approach has been extended (77) to analyze stochastic nonlinear systems, as well as input-inaccessible nonlinear systems with multiple outputs.
7. Henceforth, reference to an arbitrary system denotes a deterministic, causal, time-invariant, finite memory, continuous system.

Suppose that we can characterize an arbitrary system in terms of its response to a Gaussian white noise input, for example by using the Wiener or Volterra series. Then we have effectively identified (nonparametrically) the system because we can now theoretically derive its response to any input. Thus it might appear that the best input for identifying a nonlinear system is Gaussian white noise. However, this remarkable notion is tempered by:

1. The practical problems of actually generating true Gaussian white noise inputs (particularly in physical identification experiments). Deviations from the ideal white Gaussian input introduce estimation errors. Moreover, Victor (76) has studied how the information content of laboratory approximations of white Gaussian noise limits the ability to identify the terms of a functional expansion. He shows that the fractal dimension (specifically, the capacity dimension) of the laboratory test input signal limits the number of terms identifiable in an orthogonal functional expansion representation.
2. The numerical problems involved in obtaining the unknowns in the Wiener or Volterra series, namely the Wiener or Volterra kernels. Even with noise-free data, determining the Wiener kernels from moderate length input-output signals by cross-correlation can result in major estimation errors (obviously depending on the kernel memory length). For example for a second-order Volterra (or Wiener) series the second-order kernel estimates are typically considerably less accurate than the first-order kernel estimates. Use of lengthy inputs to improve accuracy is frequently not feasible. Aside from computation time another problem is that the system characteristics may change over the duration of a lengthy experiment (i.e., the assumption that the system is time-invariant may no longer be warranted).
3. The redundancy in Gaussian inputs arguably makes them less efficient than certain pseudorandom signals (74). For example a very interesting, elegant, and promising method for determining "Volterra-like" kernels of a system based on multilevel M-sequences (preferably inverse-repeat) has been recently proposed by Sutter (74). The method determines extremely rapidly the kernel estimates

(even high-order kernels) from locations on the first-order cross-correlation function. If the system order is high, then unless the system has a relatively short memory length, a very long input sequence must be used to avoid overlap of kernel slices. Such long input sequences cause problems when the system cannot be considered time-invariant over a lengthy period. If the system has lower order (e.g. second), then a much shorter stimulus length may suffice to avoid overlap. The technique further requires that the *M*-sequence input be exactly delivered to the system and this is not always possible. Note that the kernels recovered by the technique will in general depend upon the number of distinct input levels. For example, for a binary *M*-sequence, binary kernels are recovered whose diagonal values are indeterminable. Moreover the off-diagonal kernel values (even the lower-order ones) will, for a higher-order system, in general differ from both the Volterra and the Wiener kernel values. Moreover, neither the Wiener nor the Volterra kernels can generally be computed from the binary kernels.

4. The difference between Wiener and binary kernels may well be significant depending upon the application. For example, for an LNL cascade structure, the Wiener kernels are proportional (28,29) to the Volterra kernels provided that the latter exist. This proportionality enables a simple identification of the components of the LNL system (28–30,32). However, the simple proportionality with the Volterra kernels does not hold for the binary kernels. In addition a binary input may be ineffective in evoking the nonlinear behavior of a system. For example, if the system is the sum of a Hammerstein structure and a non-Hammerstein structure, the Hammerstein component will not be detected by a binary input. It should be noted that a system may be continuous (in that "small" changes in the system input result in small changes in the system output) but the kernels of the system need not be continuous functions. Thus, it may not be possible to infer diagonal kernel values from off-diagonal values even if the system is continuous. Kernels, with diagonal values determined, can be obtained using more input levels, for example, a quaternary signal (74), although this will require a longer test stimulus. Nevertheless the Sutter (74) technique holds great promise for analyzing nonlinear systems.
5. The Volterra series may be orthogonalized for other inputs (see [38,55]), which may be more optimal in some sense. Moreover in many situations it may only be possible to monitor the input rather than manipulate it. Consequently it may be necessary to orthogonalize the Volterra series with respect to the given input.

The problem of finding a practical method for determining the Wiener kernels proved a most difficult one to solve. Ideally one would like some device where each kernel could be adjusted until the desired output was obtained. Unfortunately no such device existed prior to the advent of digital computers. Wiener's approach was to represent the kernels using an orthogonal set of basis functions, much in the way that a periodic signal may be represented using an orthogonal set in the Fourier series. The orthogonal set Wiener suggested was the set of Laguerre functions, because these functions can be represented by a series of phase-shift electrical networks called lattice networks (83). Thus Wiener's idea was to represent the kernels using Laguerre functions that are implemented practically by individual electronic components. By electrical manipulation involving the lattice networks, the coefficients of Hermite polynomials of the Laguerre functions, and therefore the Wiener kernels, could be altered. Of course, Wiener's approach was not limited to the use of Laguerre functions for the set of basis functions.

Wiener argued that two systems might be considered equivalent if their responses to a Gaussian white noise input are minimally different in the mean-square sense. Thus suppose one of the systems is the unknown nonlinear system to be analyzed or synthesized and the other the physical model implementing the Laguerre functions and Hermite polynomials. Then analysis or synthesis is considered complete when the components representing the Hermite polynomials operating on the lattice networks are adjusted such that the minimum mean-square difference between the outputs of the two systems in response to the same Gaussian white noise input is found.

The benefits of having an orthogonal representation of these kernels (e.g., using the Hermite and Laguerre functions) are enormous. If the coefficients of Hermite polynomials of Laguerre functions, as implemented electronically, are indicated by a meter and modified by a gain control, then orthogonality ensures that manipulation of the ith gain control affects only the ith meter and does not affect readings on the others. This corresponds to the fact that each additional term added to the Wiener series does not affect the coefficients of the previous terms. Hence, if there are K gain controls, only K individual adjustments need be made. With refinements (33,82) to correct for the actual input and finite duration used in the identification experiment, the basis function approach to kernel estimation may indeed be practical. At present, the expansion of Wiener kernels using basis functions is rarely employed in practice. The inaccuracies stem from estimating the coefficients by time-averaging, which is predicated on the assumption that the system input is a white Gaussian process applied for an infinite duration. Practical methods for expanding the kernels using basis functions were suggested by Watanabe and Stark (82) and Korenberg (33). In the latter method, the coefficients in the basis function expansion can be estimated accurately by orthogonalizing for the actual input and duration of the identification experiment (which we call "exact orthogonalization"). The orthogonalization enables rapid searching for a concise subset of basis functions to approximate accurately the kernels. The coefficients obtained in both methods (33,82) are *least-square estimates*, and neither method is limited to inputs which are Gaussian, white, or lengthy. Additional major advantages of the methods (33,37,82) are ready applicability to a system with lengthy memory, and robustness in the face of heavy noise contamination.

METHODS FOR DETERMINING THE WIENER KERNELS

Lee and Schetzen (44) published a relatively simple time domain technique for estimating the Wiener kernels that caused applications of functional series identification to flourish (e.g., 46,47,48,65,66,71).

ESTIMATION OF THE WIENER KERNELS BY CROSS-CORRELATION

Again let the input $x(t)$ be zero-mean Gaussian white noise with power spectral density P. Lee and Schetzen (44) showed that the Wiener kernels could be estimated from

$$k_i(\tau_1,\ldots,\tau_i) = \frac{1}{i!P^i}\overline{\left(y(t) - \sum_{m=0}^{i-1} G_m[k_m,x]\right)x(t-\tau_1)\ldots x(t-\tau_i)} \ . \tag{14}$$

When all of the $\tau_1, \ldots, \tau_i$ are distinct (i.e., unequal) this reduces to

$$k_i(\tau_1, \ldots, \tau_i) = \frac{1}{i! P^i} \overline{y(t)x(t-\tau_1)\ldots x(t-\tau_i)} \ . \tag{15}$$

Palm and Poggio (54) point out certain fundamental problems in calculating the diagonal values of third- and higher-order continuous time kernels. However, these difficulties are easily circumvented (4) using appropriate discrete stochastic inputs. For a discrete Gaussian input, replace $y(t)$ in Eq. 15 by its discrete counterpart $y(n)$, and the factors $x(t-\tau_1), \ldots, x(t-\tau_i)$ by Grad–Hermite (i.e., generalized multidimensional Hermite) polynomials (1,86) of the discrete counterparts of these factors. This enables the Wiener kernels to be evaluated along their minor and major diagonals (2). The same method is reviewed by Goussard, Krenz, and Stark (19), and has been well known for many years. Very nearly the same idea for measuring diagonal and off-diagonal kernel values is contained in the Lee and Schetzen (44) paper (on pp. 246–247, and in particular their Eq. 34). However, they consider identification of a continuous-time system and consequently employ Dirac impulse functions in place of discrete delta functions in defining the "equivalents" of the Grad–Hermite polynomials.

Goussard (17,18) has shown that improved Wiener kernel estimates are obtainable using a stochastic approximation technique. A very efficient method for measuring the Wiener kernels in the frequency domain via the fast Fourier transform algorithm was proposed by French and Butz (14).

ADVANTAGES AND DISADVANTAGES OF THE WIENER APPROACH

1. The orthogonality of the G-functionals enables Wiener kernels to be obtained without the need to solve a set of simultaneous integral equations as would be required if Volterra kernels were to be estimated using cross-correlation.
2. Rearranging the Wiener series of a given order yields the finite-order Volterra series which best approximates the system in the least-squares sense for that order and the particular Gaussian input used (46).
3. Extending the truncated Wiener series approximation of a system to include higher-order orthogonal G-functionals does not affect the values of the G-functionals already estimated.
4. As with the Volterra series, it may require a high-order truncation of the Wiener orthogonal series to represent a given nonlinear system accurately.
5. Identifying higher-order Wiener kernels by cross-correlation is computationally unwieldy and time-consuming, and may result in only slight improvement in goodness of fit.
6. The Wiener kernel values in neurophysiological applications (46,65) normally comprise a large parameter set, although the order of nonlinearities adequate for good approximation frequently remains at or below fifth ([9 Fig. 5], [10,16,39]).
7. Over-parameterization may result in approximating the measurement-noise variance, as well as the system output variance, resulting in a less accurate model for data other than the actual input-output records used in the identification.
8. Analogous orthogonalizations can be carried out for other inputs with special

auto-correlation properties, for example constant-switching-pace symmetric random signals (46,49), pseudo-random signals based on m-sequences (49), and multilevel inputs (52). Klein and Yasui (27) have shown how the kernels in the orthogonal expansions for different stimuli are related. Krausz (42) has shown how discrete Poisson white-noise can be used for kernel identification.

9. Orthogonalization is also possible for inputs lacking special auto-correlation properties, but then expressions for higher-order orthogonal functionals may become prohibitively difficult to calculate (55). A simple method for exactly orthogonalizing the Volterra series for such an input, identifying the weightings in the orthogonal series, and then reconstructing the Volterra kernels through an efficient formula is given by Korenberg *et al.* (38). This technique also permits the diagonal kernel values to be determined accurately. A fast orthogonal algorithm (35,36) for carrying out the kernel identification is also available. In addition, the exact orthogonalization for a given input can be combined (33) with parallel cascade identification, to yield fast robust kernel estimation applicable even when the system has lengthy memory (37).
10. Gate functions ([8],[9, Appendix Eq. A6], [67,69]) of the input (or of functionals of the input) have been used to develop nonlinear system representations in which the constituent terms are orthogonal for an arbitrary input. Each gate function has the value zero "everywhere" except over a single bin, where it has value unity. The bins for different gate functions do not overlap so that a set of such functions can be constructed to form an orthogonal basis. While orthogonality does not depend here on the statistical properties of the input, the system representation is highly nonparsimonious, and basically amounts to storing the average value attained by the output when the input lies in given bins (or more generally when functionals of the input lie in given bins). Luminance-projection functions of Klein and colleagues [Appendix in (9)] provide an excellent example of gate functions that simplify interpretation of kernels, and explain a well-known psychophysical visual movement phenomenon, reverse phi motion (their Eq. A12).
11. Walsh functions have also been used as a set of orthogonal functions to expand the kernels by French and Butz (15). These authors constructed a set of kernels similar to the Wiener kernels, which, however, contained the dyadic convolution operation (13). The dyadic kernels could be measured more efficiently than the normal kernels. This efficiency must be tempered by the fact that higher order dyadic kernels than normal kernels may be required to model accurately certain systems (13).
12. A valuable alternative to Wiener's stochastic input approach is the sum-of-sinusoids technique, which can be used to obtain rapidly estimates of the Fourier transforms of the kernels, and has been successfully applied in neurophysiology (79).

ESTIMATION OF THE WIENER KERNELS BY REPEATED TOEPLITZ MATRIX INVERSION

The cross-correlation techniques suffer from the disadvantage that unless the input is almost exactly white and very long, poor estimates of the kernels may be obtained. In practice long input records are inconvenient and white inputs are almost impossible to implement physically. Schetzen (68) has given equations (resulting from cross-correlation), which contain the Wiener kernels when the input is colored Gaussian (see also the work of Bedrosian and Rice [3] and the book by Marmarelis and Marmarelis [46]). However a practical method of solving these equations in the time domain for the kernels when the input is not white (and possibly non-Gaussian) has been lacking.

A discrete-time, continuous, finite-memory, physically realizable system can be uniformly approximated to a given degree of accuracy by a discrete Volterra series of sufficient, but finite, order J:

$$y_s(n) = \sum_{j=0}^{J} \sum_{i_1=0}^{I} \ldots \sum_{i_j=0}^{I} h_j(i_1,\ldots,i_j)x(n-i_1)\ldots x(n-i_j) \ . \tag{16}$$

(In Eq. 16, the term corresponding to $j = 0$ is the constant h_0.) This approximation is uniformly convergent when x belongs to a given set of uniformly bounded signals.

Consider a second-order discrete-time finite memory Volterra series:

$$y(n) = h_0 + \sum_{i=0}^{I} h_1(i)x(n-i) + \sum_{i_1=0}^{I} \sum_{i_2=0}^{I} h_2(i_1,i_2)x(n-i_1)x(n-i_2) \ . \tag{17}$$

Assume that the input, x, is colored Gaussian. It follows that

$$\begin{aligned} \phi_{xy}(j) &\equiv \overline{y(n)x(n-j)} \\ &= \sum_{i=0}^{I} h_1(i)\overline{x(n-i)x(n-j)} \\ &= \sum_{i=0}^{I} h_1(i)\phi_{xx}(j-i) \ . \end{aligned} \tag{18}$$

Note that in the second and third lines immediately above, the term involving the second-order kernel was omitted since it vanishes under the time average. This is because once the second-order term in Eq. 17 is multiplied by $x(n-j)$, its time average is taken over an odd number of x terms and is therefore zero. (The term involving the constant h_0 was omitted for a similar reason.) Next, it can be shown that

$$\begin{aligned} \phi_{xxy}(j_1,j_2) &\equiv \overline{(y(n)-\bar{y})x(n-j_1)x(n-j_2)} \\ &= \sum_{i_1=0}^{I} \sum_{i_2=0}^{I} h_2(i_1,i_2)\overline{x(n-i_1)x(n-i_2)x(n-j_1)x(n-j_2)} \\ &\quad - \bar{y}\overline{x(n-j_1)x(n-j_2)} \\ &= 2\sum_{i_1=0}^{I} \sum_{i_2=0}^{I} h_2(i_1,i_2)\phi_{xx}(j_1-i_1)\phi_{xx}(j_2-i_2) \ . \end{aligned} \tag{19}$$

Schetzen (68) has obtained analogous equations involving the Wiener kernels, for a continuous-time system. These equations can readily be solved for the kernels in the frequency domain by division (46,68). However, we examine here a simpler solution, which is achievable in the time-domain.

Equation 18 can be solved for the first-order kernel by a linear Toeplitz matrix inversion (24). By an extension (33), Eq. 19 and analogous equations, involving higher-order kernels, can be solved by repeated Toeplitz matrix inversions very conveniently. While the method is illustrated here for identifying first- and second-order Volterra kernels (which are equivalent to the corresponding Wiener kernels for a system represented by a second-order functional series), it can be used to obtain Wiener kernels of all orders.

Equation 18 holds for $j = 0, \ldots, I$ which leads to $I + 1$ linear equations in the $I + 1$ unknowns $h_1(0), \ldots, h_1(I)$. The solution of these equations may be undertaken by inversion of a matrix $A = \langle a_{ij} \rangle$ which is Toeplitz (i.e., a_{ij} is constant for all elements having the same value of i-j) and symmetric. (Here $a_{ij} = \phi_{xx}(i - j)$.) A very efficient computer program (87) exists for inversion of such matrices and may be used to obtain the values of the first-order kernel $h_1(i)$.

To find the second-order kernel, define

$$g(j_1, i_2) = \sum_{i_1=0}^{I} h_2(i_1, i_2)\phi_{xx}(j_1 - i_1) \tag{20}$$

for $j_1, i_2 = 0, \ldots, I$. Then Eq. 19 becomes

$$\phi_{xxy}(j_1, j_2) = 2 \sum_{i_2=0}^{I} g(j_1, i_2)\phi_{xx}(j_2 - i_2) \tag{21}$$

for $j_1, j_2 = 0, \ldots, I$.

For a fixed j_1, first solve Eq. 21 for $g(j_1, i_2)$, $i_2 = 0, \ldots, I$, using Toeplitz matrix inversion. By repeating this for different values of j_1, obtain $g(j_1, i_2)$, $i_2 = 0, \ldots, I$ and $j_1 = 0, \ldots, I$. Then solve Eq. 20 for $h_2(i_1, i_2)$, $i_1 = 0, \ldots, I$ for fixed i_2 using inversion of the same Toeplitz matrix. Repeating for different values of i_2 yields the kernel $h_2(i_1, i_2)$ *for all* $i_1, i_2 = 0, \ldots, I$. To improve accuracy, it is desirable to subtract the output due to the first-order kernel, $h_1(j)$, from the system output before computing the second-order cross-correlation on the left side of Eq. 21.

The same procedure can be used analogously to identify higher-order kernels by Toeplitz matrix inversion. Since it is always the same Toeplitz matrix which is repeatedly inverted, in certain applications it may be more efficient to store the inverse. Finally, while the procedure is presented here for a Gaussian input, in practice we have obtained good results by this method for other stochastic inputs.

EXAMPLE OF SIMULATED SYSTEMS

The first simulated example was a simple Wiener model system, having a second-order functional series representation, consisting of a low-pass, under-damped, second-order, dynamic linear subsystem (Fig. 2) followed by a simple squaring static nonlinearity. The memory of the simulated Wiener model lasted for 40 sample values. It should be noted that in this example all Wiener and Volterra kernels are identically zero except for the second-order Wiener and Volterra kernels (which are equal in this case), and the zero-order Wiener kernel (which equals the output mean). The

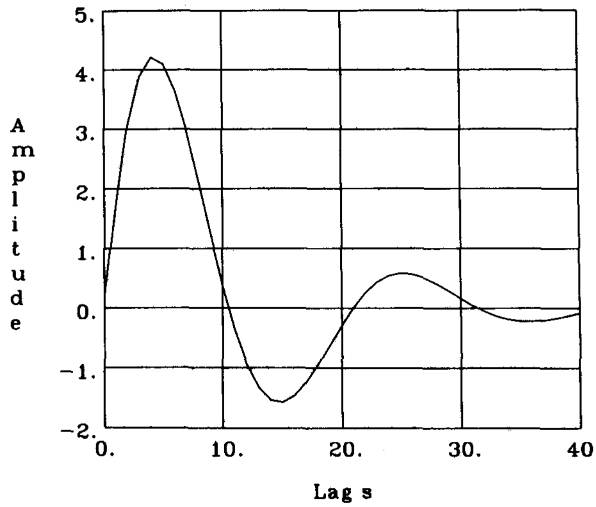

FIGURE 2. The dynamic linear subsystem preceding the static nonlinearity in our simulated Wiener model used to illustrate kernel identification by repeated Toeplitz matrix inversion and by classical cross-correlation. The dynamic linear subsystem is a second-order low-pass underdamped system and the static nonlinear element is a simple squarer.

first- and second-order kernels were estimated using both the technique outlined above and the traditional cross-correlation approach for white Gaussian inputs.

The input was a 10,000 sample Gaussian colored noise signal generated using

$$x(n) = w(n) - 0.5w(n-1) - 0.5w(n-2)$$

where w is zero-mean, unity variance, white Gaussian noise produced using the technique proposed by Hunter and Kearney (23). The auto-correlation function of the input, x, is shown in Fig. 3 where it can be seen that the degree of coloring is not pronounced. Again the simulations in this example were carried out in the NEXUS language for simulation, system and signal analysis (25).

A negligible first-order Wiener (or Volterra) kernel was estimated for the simulated system. The estimation involved a linear analysis using Toeplitz matrix inversion (24). The output variance accounted for using this estimate was essentially zero (actually 0.5%), as expected since the true first-order kernel is identically zero.

Figure 4 shows the true second-order Wiener (or Volterra) kernel of the simulated system. The second-order cross-correlation estimate of the second-order kernel is shown in Fig. 5. The estimate involved determining the second-order cross-correlation after subtracting from the system output both the output mean and the output of the estimated first-order kernel (a very similar estimate was obtained when the mean alone was subtracted). [If the input had been white this would have been an application of the Lee–Schetzen (44) kernel estimation technique. While these authors specified that the input be *white* Gaussian, other researchers might be tempted to discount the small amount of coloring of our input (as evidenced in Figure 3), and attempt kernel estimation by cross-correlation.] The poor approximation of this kernel

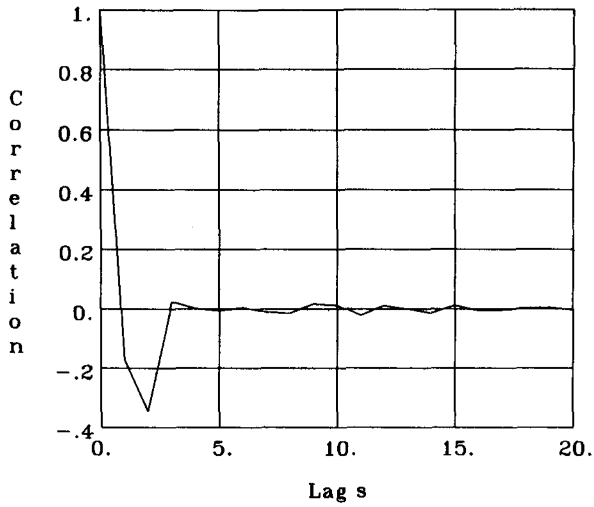

FIGURE 3. Autocorrelation function of the colored Gaussian input applied to the simulated Wiener model system used to illustrate the kernel identification procedures.

estimate is due both to the finite record length and the failure to correct for the non-white input. Smoothing this estimate did not improve the approximation significantly.

Figure 6 shows the second-order Wiener (or Volterra) kernel of the system estimated by the repeated Toeplitz matrix inversion approach (33) described above. Comparison of Figs. 5 and 6 with Fig. 4 illustrates the improvement afforded by use of a very fast technique compensating for nonwhite inputs. The advantage of this method is particularly relevant since it is usually impossible to probe physical nonlinear system

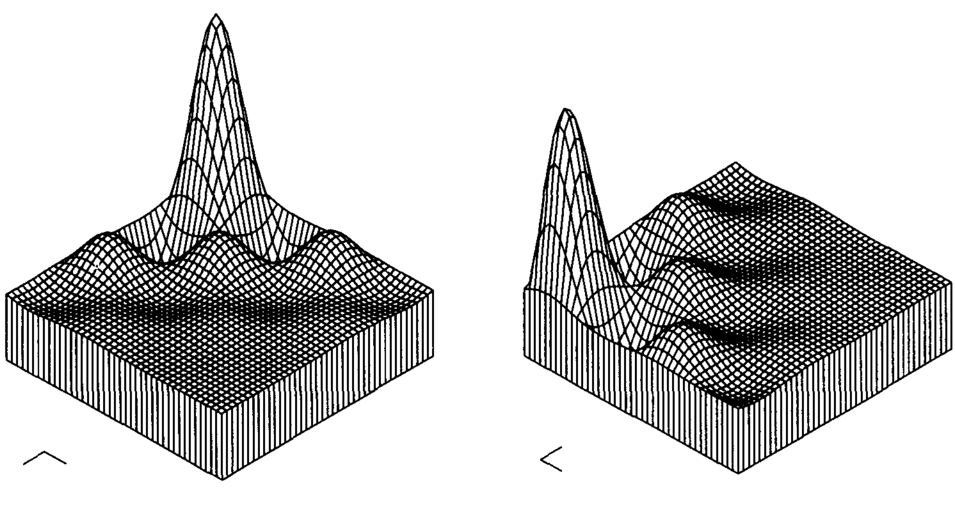

FIGURE 4. The true second-order Wiener (or Volterra) kernel of the simulated Wiener model system used to illustrate the kernel identification procedures.

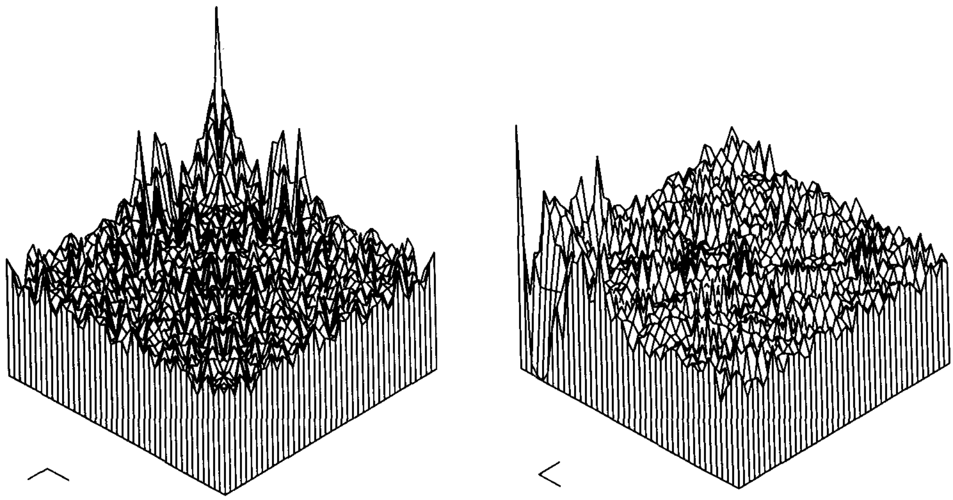

FIGURE 5. Estimated second-order Wiener (or Volterra) kernel, of the system whose true kernel is shown in Fig. 4, obtained by second-order cross-correlation after subtracting from the system output both the output mean and the output of the estimated first-order kernel. (If the input had been white this would have been an application of the Lee–Schetzen kernel estimation technique.)

with white inputs due to the inherent frequency response limitations of most input generators.

For comparison, Fig. 7 shows the first- and second-order Wiener kernels, estimated by the method (33,37) which combines exact orthogonalization for the given input with basis function expansion of the kernels. Here the simulated (i.e., true) system

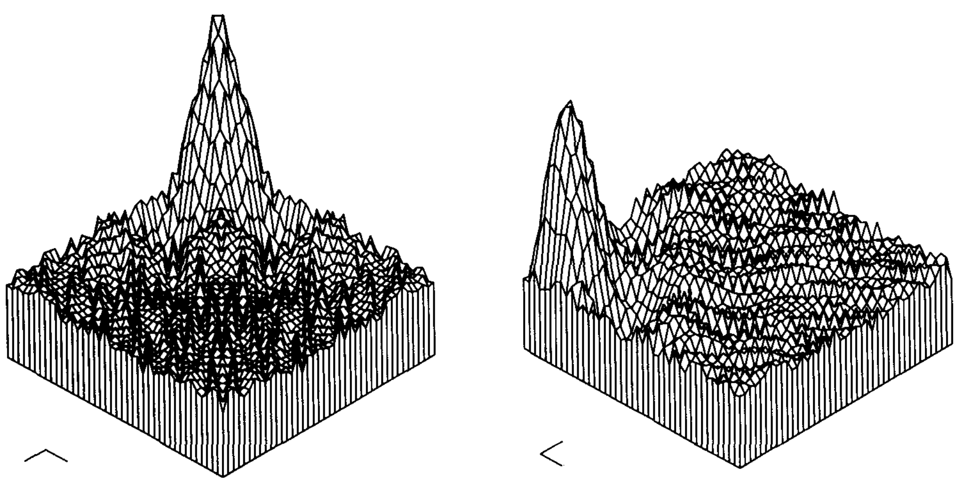

FIGURE 6. Estimated second-order Wiener (or Volterra) kernel, of the system whose true kernel is shown in Fig. 4, calculated by the repeated Toeplitz matrix inversion approach.

was again second-order, and the basis functions employed were 14 decaying exponentials (37), the "white" Gaussian input length was 500 samples, and the memory length was 25 samples. As shown in Figure 7a the first-order kernel estimates (points) were virtually coincident with the true kernel values (solid line). The true second-order Wiener kernel (Figure 7b) was estimated (Figure 7c) with slightly more error, but still with far greater accuracy than obtainable by the Lee–Schetzen cross-correlation method (44) on such a short record. Running time for the estimation was comparable to that for the Toeplitz method for an equivalent number of points (ie., very fast). It should be noted that the good performance with the exponential basis functions shown here was not always observed in other applications, and frequently sinusoids, exponentially decaying sinusoids, or the discrete form of the Laguerre functions favoured by Wiener provide a more economical approximation of the kernels (and hence more accuracy for a given number of basis functions). A refinement over Wiener's basis function approach is provided by the exact orthogonalization (33). This substantially eliminates the errors due to finite record length and deviations from a white Gaussian input, which plagued practical implementations of Wiener's formulation. Exact orthogonalization also enables the potential benefit of selecting any particular basis function to be assessed rapidly. Note that least-squares estimates (determined over the record length) could also be directly obtained of the coefficients of the basis function expansion of the kernels using the method of Watanabe and Stark (82).

USE OF THE WIENER FUNCTIONAL SERIES APPROACH TO INFER MODEL STRUCTURE

The Wiener functional series approach appears to be most useful when little a priori knowledge of the system under study is available. Inspection of the kernels (or the equivalent cross-correlations) may lead to insight into the structural form of an appropriate model for the nonlinear system. In suitable cases this enables subsequent use of a block structured or cascade model approach.

All of the following tests for cascade structure are necessary but not sufficient conditions (unless further assumptions are made about the structure of the system). For a dynamic linear subsystem followed by a static nonlinearity, called a Wiener cascade model (not to be confused with the Wiener functional series), it is necessary (26) in the discrete time domain that

$$\phi_{xxy}(j_1,j_2) = C\phi_{xy}(j_1)\phi_{xy}(j_2)$$

for all j_1 and j_2, where C is a constant, if $\phi_{xy} \neq 0$. Equivalently in the discrete frequency-domain (29)

$$\Phi_{xxy}(\omega_1,\omega_2) = C\Phi_{xy}(\omega_1)\Phi_{xy}(\omega_2)$$

for all ω_1 and ω_2.

For a static nonlinearity followed by a dynamic linear subsystem, called a Hammerstein cascade model, it is necessary (26) in the discrete time domain that

$$\phi_{xxy}(j_1,j_2) = D\phi_{xy}(j_1)\delta(j_1 - j_2)$$

for all j_1 and j_2, *where* D is a constant, if $\phi_{xy} \neq 0$. Here the discrete delta function $\delta(j) = 0, j \neq 0$ *and* $\delta(0) = 1$. Equivalently in the discrete frequency-domain (29)

$$\Phi_{xxy}(\omega_1, \omega_2) = D\Phi_{xy}(\omega_1 + \omega_2)$$

for all ω_1 and ω_2.

For a system to have the LNL structure of two dynamic linear systems "sandwiching" a static nonlinearity, it is necessary (40) that

$$\sum_{j=0}^{\infty} \phi_{xxy}(i,j) = E\phi_{xy}(i)$$

where E is a constant, if $\phi_{xy} \neq 0$. Equivalently, in the discrete frequency-domain (29)

$$\Phi_{xxy}(\omega,0) = E\Phi_{xy}(\omega) \ .$$

Moreover, the shape of the first dynamic linear system in the LNL cascade is given (30,32) by the first non-negligible term of the sequence $\phi_{xxy}(j,0), \phi_{xxy}(j,1), \ldots$.

Necessary conditions for more elaborate block structures are given by Haber (21).

APPLICATIONS OF THE WIENER KERNEL APPROACH

The Wiener kernel representation has probably been most extensively used in biology and physiology (see reviews in [46,65]). The work of Stark and coworkers (66,71) in measuring the Wiener kernels of the pupillary control system has already been cited above. By examining the first- and second- order Wiener kernels, Stark (72)

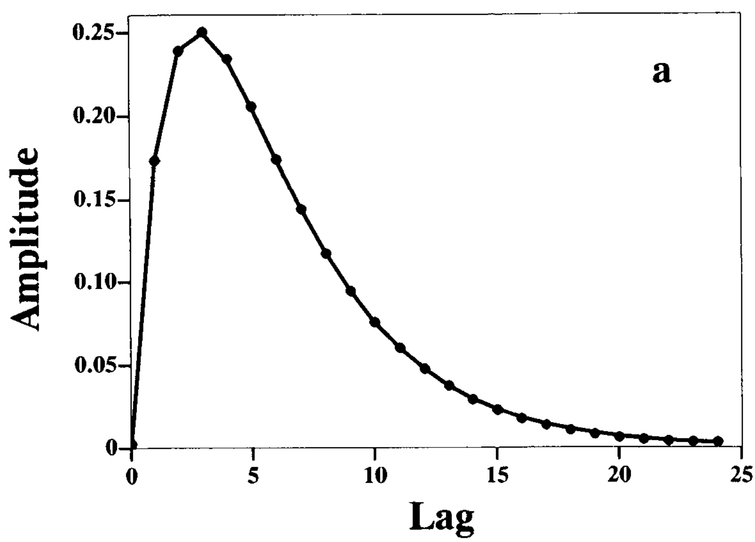

FIGURE 7. Estimation of kernels via the method which combines basis function expansion with exact orthogonalization. a. Estimated first-order kernel (points), superimposed on the true first-order kernel (solid line). *(Figure continued on facing page.)*

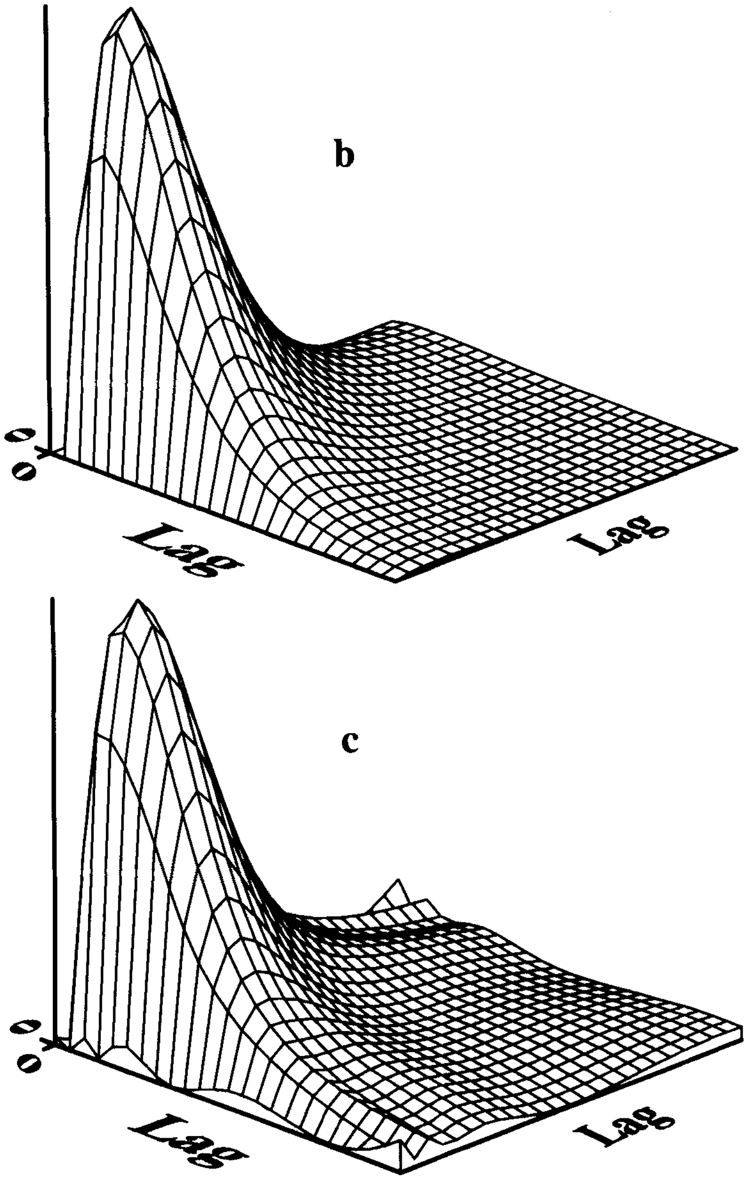

FIGURE 7 continued. Estimation of kernels via the method which combines basis function expansion with exact orthogonalization. b. true second-order kernel; c. estimated second-order kernel.

inferred that the system could be approximated by a Hammerstein cascade model. This inference resulted from the dominance of values of the second-order kernel main diagonal (over off-diagonal elements) which closely resembled a scaled version of the first-order kernel.

An early use of Wiener functional analysis was made by Marmarelis and Naka (47,48) in neural process modeling. They investigated a three-stage neural chain in the catfish retina, comprising horizontal, bipolar, and ganglion cells. Preliminary analysis indicated that this system could be approximated with little error by a second-order functional expansion. Marmarelis and Naka (47) injected an approximately white-Gaussian intra-cellular current stimulus and recorded the output neural pulse train at the ganglion cell stage. By repeatedly applying the same stimulus they generated a poststimulus time histogram (which could be treated as a continuous output) of ganglion cell firing. They then identified the first- and second-order Wiener kernels relating the input current to this output. (Subsequently, Sakuranaga *et al.* [64] showed that repetition of the stimulus was unnecessary. Rather the kernels could be measured by cross-correlating the input with a "unitary" output, whose value was one at the time of an action potential and zero at all other times.) By examining these kernels Marmarelis and Naka inferred that the neural chain could be approximated mathematically by a Wiener cascade structure (in their case, a dynamic linear subsystem in the form of a low-pass differentiator followed by a half-wave rectifier). This represented an early attempt to use the functional series approach to infer a block structured model (26,40).

The innovative application of the white noise approach by Naka and colleagues (e.g., the work of Sakai and Naka [57–62]) has led to major advances in our understanding of signal processing in the outer and inner vertebrate retina. The strong similarity and slight discrepancy between light-evoked slow potential and spike kernels (57,64) has recently been analyzed and an explanation provided (41). In addition, Marmarelis and McCann (50) used the Wiener kernel approach to model the photoreceptor of the fly *Calliphora Erythrocephala*. By inspection of the kernels they inferred that the physiological system studied might be approximated by the Wiener cascade model. For a brief, recent review of white noise applications in visual neuroscience, see reference (63).

Hunter (22) studied the relation between frog single active muscle fiber stochastic length perturbations (input) and resulting muscle tension fluctuations (output). The Wiener model was found to provide an excellent fit to the experimental data, in contrast to the Hammerstein model. Moreover the use of the nonlinear Wiener model avoided the need for input-dependent time-constants required when the input-output was modeled using a linear system alone.

Korenberg, French and Voo (39), and French and Korenberg (16), have identified an LNL model for neural encoding by an insect mechano-receptor, using a recent accurate technique (35) to estimate the Wiener kernels. An LNL model has been applied successfully to as complex a structure as the cat's visual cortex (11).

The orthogonal approach introduced by Wiener (83) for nonlinear system identification has been extended and applied in many other areas. For example, a fast orthogonal search method (34,35,36) has been developed to fit time-series data by a parsimonious sum of sinusoids which are not necessarily commensurate. This method is capable of eight times finer frequency resolution than attainable by a conventional Fourier series analysis.

CONCLUSION

The Wiener functional series approach has been used extensively in the analysis of nonlinear biological and physiological systems. Its appeal lies in the ability to model the mathematical relation between system input and output with no a priori knowledge of system structure. A concomitant disadvantage is that the resulting model may merely mimic the mathematical operation of the system under study without affording insight into the underlying mechanism by which it is carried out. However on certain occasions (e.g., in pattern classification using nonlinear discriminant functions) all that is required of the nonlinear representation is the ability to predict the system output. In other instances Volterra and Wiener kernel analyses have led to inferences of model structure.

A very efficient time-domain identification of Wiener kernels when the input is colored was illustrated here on a simple example of a Wiener model structure. The identification method utilizes repeated Toeplitz matrix inversion to arrive rapidly at kernel estimates that are significantly improved over those obtained by cross-correlation, even when the input is only lightly colored. We have also pointed out an error frequently made in using two proposed tests of nonlinearity of a system. In particular, we have shown that there are infinitely many (nonlinear) Wiener model structures which satisfy the putative "sufficient" conditions for linearity which have sometimes been erroneously read into the tests.

REFERENCES

1. Barrett, J.F. The use of functionals in the analysis of nonlinear physical systems. J. Elect. Control. 15:567–615; 1963.
2. Barrett, J.F. Functional series representation of nonlinear systems—some theoretical comments. IFAC Symp. Ident. Sys. Param. Est. 1:251–256; 1982.
3. Bedrosian, E.; Rice, S.O. The output properties of Volterra systems (nonlinear systems with memory) driven by harmonic and Gaussian inputs. Proc. IEEE 59:1688–1707; 1971.
4. Billings, S.A. An overview of nonlinear systems identification. IFAC Symp. Ident. Sys. Param. Est. 1:725–729; 1985.
5. Billings, S.A.; Fakhouri, S.Y. Identification of systems containing linear dynamic and static nonlinear elements. Automatica. 18:15–26; 1982.
6. Billings, S.A.; Voon, W.S.F. Structure detection and model validity tests in the identification of nonlinear systems. IEE Proc. 130, Part D:193–199; 1983.
7. Billings, S.A.; Voon, W.S.F. Correlation based model validity tests for non-linear models. Int. J. Control. 44:235–244; 1986.
8. Bose, A.G. A theory of nonlinear systems. MIT Res. Lab. Elec. Tech. Rep. No. 309, 1956.
9. Emerson, R.C.; Citron, M.C.; Vaughn, W.J.; Klein, S.A. Nonlinear directionally selective subunits in complex cells of cat striate cortex. J. Neurophysiol. 58:33–65; 1987.
10. Emerson, R.C.; Korenberg, M.J.; Citron, M.C. Measurement of a simple-cell threshold function in cat's striate cortex. Soc. Neurosci. Abstr. 14:899; 1988.
11. Emerson, R.C.; Korenberg, M.J.; Citron, M.C. Identification of intensive nonlinearities in cascade models of visual cortex and its relation to cell classification. In: Marmarelis, V.Z., ed. Advanced methods of physiological system modeling, Vol. 2, New York: Plenum Press; 1989: pp. 97–111.
12. Frechet, M. Sur les fonctionnelles continues. Annales Scientifiques de l'Ecole Normal Superieure. 27:193–219; 1910.
13. French, A.S. Measuring the Wiener kernels of a nonlinear system by use of the fast Fourier transform and Walsh functions. Proc. 1st Symp. Testing Ident. Nonlin. Sys. 1:76–88; 1975.
14. French, A.S.; Butz, E.G. Measuring the Wiener kernels of a nonlinear system using the fast Fourier transform algorithm. Int. J. Contr. 17:529–539; 1973.
15. French, A.S.; Butz, E.G. The use of Walsh functions in the Wiener analysis of nonlinear systems. IEEE Trans. Comp. 23:225–232; 1974.
16. French, A.S.; Korenberg, M.J. A nonlinear cascade model for action potential encoding in an insect sensory neuron. Biophys. J. 55:655–661; 1989.
17. Goussard, Y. Identification de systemes nonlineaires, representes par developpements fonctionnels et soumis a diverses entree aleatoires, a l'aide d'une methode d'approximation stochastique. These de Docteur-Ingenieur, Universite de Paris-Sud, Centre d'Orsay; 1983.
18. Goussard, Y. Wiener kernel estimation: A comparison of cross-correlation and stochastic approxima-

tion methods. In: Marmarelis, V.Z., ed. Advanced methods of physiological system modeling. Los Angeles: USC Biomedical Simulations Resource, vol. 1; 1987: pp. 289–302.

19. Goussard, Y.; Krenz, W.C.; Stark, L. An improvement of the Lee and Schetzen cross-correlation method. IEEE Trans. Automat. Contr. 30:895–898; 1985.
20. Haber, R. Nonlinearity tests for dynamic processes. IFAC Symp. Ident. Sys. Param. Est. 1:409–413; 1985.
21. Haber, R. Structure identification of block-oriented models based on the Volterra kernels. IFAC Symp. Ident. Sys. Param. Est. (Appendix); 1985.
22. Hunter, I.W. Experimental comparison of Wiener and Hammerstein cascade models of frog muscle fiber mechanics. Biophys. J. 49:81a; 1986.
23. Hunter, I.W.; Kearney, R.E. Generation of random sequences with jointly specified probability density and autocorrelation functions. Biol. Cybern. 47:141–146; 1983.
24. Hunter, I.W.; Kearney, R.E. Two-sided linear filter identification. Med. Biol. Eng. Comput. 21:203–209; 1983.
25. Hunter, I.W.; Kearney, R.E. NEXUS: A computer language for physiological systems and signal analysis. Comp. Biol. Med. 14:385–401; 1984.
26. Hunter, I.W.; Korenberg, M.J. The identification of nonlinear biological systems: Wiener and Hammerstein cascade models. Biol. Cybern. 55:135–144; 1986.
27. Klein, S.; Yasui, S. Nonlinear systems analysis with non-Gaussian white stimuli: General basis functionals and kernels. IEEE Trans. Inform. Theory. 25:495–500; 1979.
28. Korenberg, M.J. Cross-correlation analysis of neural cascades. Proc. 10th Annual Rocky Mountain Bioeng. Symp., 47–52; 1973.
29. Korenberg, M.J. Identification of biological cascades of linear and static nonlinear systems. Proc. 16th Midwest Symp. Circuit Theory. 18.2:1–9; 1973.
30. Korenberg, M.J. Statistical identification of parallel cascades of linear and nonlinear systems. IFAC Symp. Ident. Sys. Param. Est. 1:580–585; 1982.
31. Korenberg, M.J. Orthogonal identification of nonlinear difference equation models. Midwest Symp. Circuit Sys. 1:90–95; 1985.
32. Korenberg, M.J. Identifying noisy cascades of linear and static nonlinear systems. IFAC Symp. Ident. Sys. Param. Est. 1:421–426; 1985.
33. Korenberg, M.J. Functional expansions, parallel cascades and nonlinear difference equations. In: Marmarelis, V.Z. ed. Advanced methods of physiological system modeling. Los Angeles: USC Biomedical Simulations Resource, vol 1; 1987:pp. 221–240.
34. Korenberg, M.J. Fast orthogonal identification of nonlinear difference equation and functional expansion models. Midwest Symp. Circuit Sys. 1:270–276; 1987.
35. Korenberg, M.J. Identifying nonlinear difference equation and functional expansion representations: The fast orthogonal algorithm. Ann. Biomed. Eng. 16:123–142; 1988.
36. Korenberg, M.J. A robust orthogonal algorithm for system identification and time-series analysis. Biol. Cybern. 60:267–276; 1989.
37. Korenberg, M.J. A rapid and accurate method for estimating the kernels of a nonlinear system with lengthy memory. 15th Biennial Symp. Commun. June, 1990; Queen's University, Kingston, Canada, pp. 57–60.
38. Korenberg, M.J.; Bruder, S.B.; McIlroy, P.J. Exact orthogonal kernel estimation from finite data records: Extending Wiener's identification of nonlinear systems. Ann. Biomed. Eng. 16:201–214; 1988.
39. Korenberg, M.J.; French, A.S.; Voo, S.K.L. White-noise analysis of nonlinear behavior in an insect sensory neuron: Kernel and cascade approaches. Biol. Cybern. 58:313–320; 1988.
40. Korenberg, M.J.; Hunter, I.W. The identification of nonlinear biological systems: LNL cascade models. Biol. Cybern. 55:125–134; 1986.
41. Korenberg, M.J.; Sakai, H.M.; Naka, K-I. Dissection of the neuron network in the catfish inner retina. III. Interpretation of spike kernels. J. Neurophysiol. 61:1110–1120; 1989.
42. Krausz, H.I. Identification of non-linear systems using random impulse train inputs. Biol. Cybern. 19:217–230; 1975.
43. Lee, Y.W. Contribution of Norbert Wiener to linear theory and nonlinear theory in engineering. In Selected Papers of Norbert Wiener, SIAM, MIT Press, 17–33; 1964.
44. Lee, Y.W.; Schetzen, M. Measurement of the Wiener kernels of a non-linear system by cross-correlation. Int. J. Control. 2:237–254; 1965.
45. Leontaritis, I.J.; Billings, S.A. Input-output parametric models for nonlinear systems; Part I—Deterministic nonlinear systems; Part II—Stochastic nonlinear systems. Int. J. Control. 41:303–359; 1985.

46. Marmarelis, P.Z.; Marmarelis, V.Z. Analysis of physiological systems. New York: Plenum Press; 1978.
47. Marmarelis, P.Z.; Naka, K.I. White-noise analysis of a neuron chain: An application of the Wiener theory. Science. 175:1276–1278; 1972.
48. Marmarelis, P.Z.; Naka, K.I. Nonlinear analysis and synthesis of receptive-field responses in the catfish retina. J. Neurophys. 36:605–648; 1973.
49. Marmarelis, V.Z. Identification of nonlinear systems through quasi-white test signals. Pasadena California: Calif. Inst. Tech.; 1976. Ph.D. Thesis.
50. Marmarelis, V.Z.; McCann, G.D. A family of quasi-white random signals and its optimal use in biological system identification. Part II: Application to the photoreceptor of Calliphora Erythrocephala. Biol. Cybern. 27:57–62; 1977.
51. McIlroy, P.J.H. Applications of nonlinear systems identification. Kingston, Ontario, Canada: Queen's University; 1986. MSc Thesis.
52. Melton, R.B.; Kroeker, J.P. Wiener functional for an N-level uniformly distributed discrete random process. IFAC Symp. Ident. Sys. Param. Est. 2:1169–1174; 1982.
53. Palm, G. On the representation and approximation of nonlinear systems. Part II: Discrete time. Biol. Cybern. 34:49–52; 1979.
54. Palm, G.; Poggio, T. The Volterra representation and the Wiener expansion: Validity and pitfalls. SIAM J. Appl. Math. 33:195–216; 1977.
55. Palm, G.; Poggio, T. Stochastic identification methods for nonlinear systems: an extension of the Wiener theory, SIAM J. Appl. Math. 34:524–534; 1978.
56. Rugh, W.J. Nonlinear system theory: The Volterra/Wiener approach. Johns Hopkins Univ Press, Baltimore; 1981.
57. Sakai, H.M.; Naka, K-I. Signal transmission in the catfish retina. IV transmission to ganglion cells. J. Neurophysiol. 58:1307–1328; 1987.
58. Sakai, H.M.; Naka, K-I. Signal transmission in the catfish retina. V. Sensitivity and circuit. J. Neurophysiol. 58:1329–1350; 1987.
59. Sakai, H.M.; Naka, K-I. Dissection of the neuron network in the catfish inner retina. I. Transmission to ganglion cells. J. Neurophysiol. 60:1549–1567; 1988.
60. Sakai, H.M.; Naka, K-I. Dissection of the neuron network in the catfish inner retina. II. Interactions between ganglion cells. J. Neurophysiol. 60:1568–1583; 1988.
61. Sakai, H.M.; Naka, K-I. Dissection of the neuron network in the catfish inner retina. IV. Bidirectional interactions between amacrine and ganglion cells. J. Neurophysiol. 63:105–119; 1990.
62. Sakai, H.M.; Naka, K-I. Dissection of the neuron network in the catfish inner retina. V. Interactions between NA and NB amacrine cells. J. Neurophysiol. 63:120–130; 1990.
63. Sakai, H.M.; Naka, K-I.; Korenberg, M.J. White noise analysis in visual neuroscience. Vis. Neurosci. 1:287–296; 1988.
64. Sakuranaga, M.; Ando, Y-J.; Naka, K-I. Dynamics of ganglion cell response in the catfish and frog retinas. J. Gen. Physiol. 90:229–259; 1987.
65. Sakuranaga, M.; Sato, S.; Hida, E.; Naka, K-I. Nonlinear analysis: Mathematical theory and biological applications. CRC Crit. Rev. Biomed. Eng. 14:127–184; 1986.
66. Sandberg, A.; Stark, L. Wiener G-function analysis as an approach to nonlinear characteristics of human pupil light reflex. Brain Res. 11:194–211; 1968.
67. Schetzen, M. Determination of optimum nonlinear systems for generalized error criteria based on the use of gate functions. IEEE Trans. Inform. Theory. 11:117–125; 1965.
68. Schetzen, M. A theory of nonlinear system identification. Int. J. Control. 20:557–592; 1974.
69. Schetzen, M. The Volterra and Wiener theories of nonlinear systems. New York: Wiley; 1980.
70. Schetzen, M. Nonlinear system modeling based on the Wiener theory, Proc. IEEE. 69:1557–1573; 1981.
71. Stark, L.W. The pupillary control system: its nonlinear adaptive and stochastic engineering design characteristics. Automatica. 5:655–676; 1969.
72. Stark, L.W. The pupil as a paradigm for neurological control systems. IEEE Trans. Biomed. Eng. 31:919–924; 1984.
73. Strang, G. Linear algebra and its applications. (2nd Ed.) New York: Academic Press; 1980.
74. Sutter, E. A practical nonstochastic approach to nonlinear time-domain analysis. In: Marmarelis, V.Z., ed. Advanced methods of physiological system modeling. Los Angeles; USC Biomedical Simulations Resource, vol. 1. 1987:pp.303–315.
75. Varlaki, P.; Terdik, G.; Lototsky, V.A. Tests for linearity and bilinearity of dynamic systems. IFAC Symp. Ident. Sys. Param. Est. 1:427–432; 1985.

76. Victor, J.D. The fractal dimension of a test signal: implications for system identification procedures. Biol. Cybern. 57:421–426; 1987.
77. Victor, J.D.; Johannesma, P. Maximum-entropy approximations of stochastic nonlinear transducers: An extension of the Wiener theory. Biol. Cybern. 54:289–300; 1986.
78. Victor, J.D.; Knight, B. Nonlinear analysis with an arbitrary stimulus ensemble. Quart. Appl. Math. 37:113–136; 1979.
79. Victor, J.D.; Shapley, R. The nonlinear pathway of Y ganglion cells in the cat retina. J. Gen. Physiol. 74:671–687; 1979.
80. Volterra, V. Sur les fonctions qui dependent d'autres fonction. Comptes Rendus Ac. des Sciences. 142:691–695; 1906.
81. Volterra, V. Theory of functionals and of integral and integro-differential equations. New York: Dover; 1959.
82. Watanabe, A.; Stark, L. Kernel method for nonlinear analysis: Identification of a biological control system. Math. Biosci. 27:99–108; 1975.
83. Wiener, N. Nonlinear problems in random theory. Massachusetts: MIT Press; 1958.
84. Yasui, S. Stochastic functional Fourier series, Volterra series, and nonlinear systems analysis. IEEE Trans. Automat. Contr. 24:230–242; 1979.
85. Yasui, S. Wiener-like Fourier kernels for nonlinear system identification and synthesis (nonanalytic cascade, bilinear, and feedback cases). IEEE Trans. Automat. Contr. 27:677–685; 1982.
86. Zadeh, L. On the representation of nonlinear operators. IRE Wescon Conv. Rec. 2:105–113; 1957.
87. Zohar, S. Fortran subroutines for the solution of Toeplitz sets of linear equations. IEEE Trans. Acoust. Speech. Sig. Proc. 27:656–658; 1979.