Modeling of Neural Systems by Use of Neuronal Modes

Vasilis Z. Marmarelis and Melissa E. Orme

Abstract— **A methodology for modeling spike-output neural systems from input-output data is proposed, which makes use of "neuronal modes" (NM) and "multi-input threshold" (MT) operators. The modeling concept of NM's was introduced in a previously published paper [4] in order to provide concise and general mathematical representations of the nonlinear dynamics involved in signal transformation and coding by a class of neural systems. This paper presents and demonstrates (with computer simulations) a method by which the NM's are determined using the 1st- and 2nd-order kernel estimates of the system, obtained from input-output data. The MT operator (i.e., a binary operator with multiple real-valued operands which are the outputs of the NM's) possesses an intrinsic refractory mechanism and generates the sequence of output spikes. The spike-generating characteristics of the MT operator are determined by the "trigger regions" defined on the basis of data. This approach is offered as a reasonable compromise between modeling complexity and prediction accuracy, which may provide a common methodological framework for modeling a certain class of neural systems.**

I. Introduction

THE subject of mathematical modeling of neuronal function has been attracting increased attention as the quantitative means for understanding information processing and coding in the nervous system. Modeling efforts are made from the subcellular and cellular level to the integrated levels of neuronal networks and behavioral neuroscience (see, for instance, the classic textbook [1]). Specialized techniques for neural modeling include parametric and nonparametric methods, i.e., the use of differential or integral equations, respectively (see, for instance, two research volumes [2], [3] edited recently on these approaches). The great diversity of modeling approaches is a natural consequence of the immense variety of neural systems and the diverse requirements of different applications. In this context, the approach presented herein must be viewed as a modeling tool suitable only for certain classes of problems (i.e., those that involve the nonlinear dynamic relation between a discernible and observable stimulus signal and its corresponding spike-train response), as specified in Appendix I and in [4]–[6]. Note that this approach cannot be used to model the spontaneous activity of neurons or autonomous neural systems (e.g., pacemaker neurons, chaotic or unforced oscillators, etc.) since those cases lack an observable stimulus.

In a previously published paper [4], the modeling concept of "neuronal modes" (NM) was introduced in order to provide concise and general mathematical representations of the nonlinear dynamics involved in signal transformation and coding by certain neural systems (see Appendix I). This modeling concept is based on the judicious selection of the principal dynamic modes of a neural system to achieve a reasonable balance between mathematical complexity and neurophysiological relevance. The proposed modeling approach decomposes into two operational steps the nonlinear dynamics involved in the transformation of the graded stimulus or potential input signal into a sequence of output spikes. Note that it may often be physiologically appropriate to consider the graded input of the system as the output of a compressive nonlinearity (receptor or post-synaptic potentials). The first operational step involves the aforementioned NM's that perform all linear dynamic (filtering) operations on the input signal. The second operational step employs a multi-input threshold (MT) operator (i.e., a binary operator with multiple real-valued operands which are the outputs of the NM's) to produce the sequence of output spikes. Note that the MT includes a refractory mechanism which provides the model with a refractory period in the generation of spikes. The two operational steps involving NM's and MT are depicted in the block-structured model of Fig. 1. This model form was developed on the basis of the general Volterra–Wiener nonlinear system theory [6], [7], adapted to the case of spike-output systems [5], and it yields concise representations of the system dynamics when the number of NM's suitable for a specific neural system is small (for a brief overview, see Appendix I).

The employed NM's are properly defined filters which capture the important dynamic characteristics of the system. These dynamics reflect the *integrated* effect of all axodendritic and axosomatic synaptic inputs (including conduction effects) on the formation of the transmembrane potential at the axon hillock preceding the generation of an action potential (or transduction dynamics and conduction effects for a sensory system). The employed MT represents all nonlinear characteristics of the system, which incorporate any nonlinearities involved in the creation of the aforementioned composite potential at the axon hillock as well as the generation of the action potential [4]. It is critical to note that the formal decomposition of a nonlinear dynamic system into a linear

Manuscript received July 23, 1992; revised May 17, 1993. This work was supported by Grant No. RR-01861 awarded to the Biomedical Simulations Resource at USC from the National Center for Research Resources of the National Institutes of Health.

V. Z. Marmarelis is with the Departments of Biomedical and Electrical Engineering, USC, Los Angeles, CA 90089.

M. E. Orme is with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92717.

IEEE Log Number 9213155.

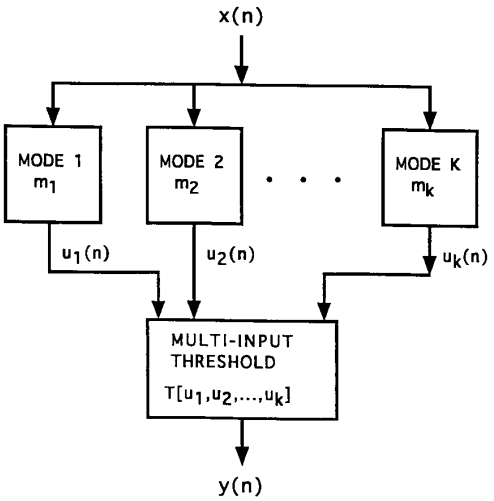


Fig. 1. Block-structured model of neural system with K neuronal modes which form a linear filter-bank representing the important dynamic characteristics of the system. The multi-input threshold operator incorporates all system nonlinearities and generates the output spikes (see Appendix I).

filter-bank (representing the dynamics) and a multi-input static nonlinearity is valid for all systems of the Wiener class (see Appendix I). The immense variety of neuronal characteristics implies a commensurate variety of NM's and MT's appropriate in each case satisfying the conditions outlined in Appendix I.

Consider, for instance, a neural system whose response depends on the amplitude and the rate of change of the stimulus signal. This can be modeled as a system with two dynamic modes corresponding to amplification and differentiation (over some finite bandwidth). Depending on the relative importance of the two modes, the system becomes more amplitude-sensitive or more rate-sensitive (e.g., tonic vs. phasic muscle spindle sensory axons, or sustained vs. transient retinal ganglion cells) and it can be concisely modeled by use of two properly weighted modes, whose outputs may be subject to nonlinear transformations prior to the generation of spikes (itself a nonlinear operation). Both of these nonlinear operations are incorporated in the MT. For instance, if the system response (spike firing frequency) depends on the size but not on the sign of the rate of stimulus change (i.e., full-wave rectification of the output of the differentiation mode seen, for example, in the "on-off transient" class of ganglion cells in the retina), then a simple MT can capture the combined nonlinear effect of rectification and spike generation [4]. Furthermore, this model can be estimated from input-output data. A variety of functional characteristics can be modeled in this framework, as long as the practical selection of appropriate NM's and MT's from input-output data is possible. This latter issue is the subject of this paper, which presents a practical methodology for this purpose and demonstrates its use by computer-simulated examples. Sections II and III present procedures for the estimation of NM's and MT's from input–output data, respectively. Section IV presents some illustrative examples using computer simulations, and Section V summarizes the main conclusions of this study.

II. Identification of Neuronal Modes

The proposed approach is based on the Volterra–Wiener theory of nonlinear system modeling and its various extensions of the last 20 years. The fundamental mathematical background concerning this paper is summarized in Appendix I. It has been shown that the model of Fig. 1 can represent the input-output relation of a broad class of spike-output systems [5], and it may find useful application to those systems that process information by use of a few principal dynamic modes [4]. This general model can be extended to systems with multiple inputs and multiple outputs (including spatio-temporal inputs/outputs relevant to vision).

The key practical issue in developing the model of Fig. 1 is how to determine the minimum number of NM filters required in a given application. We postulate that for many neural systems, the prevalent dynamic characteristics can be captured by a small number of such filters (principal dynamic modes). The notion of principal modes has been widely used in many fields of engineering and science. Consider, for instance, a linear time-invariant system. When its impulse response function is decomposed (expanded) on a general basis, then a large number of components generally results. Nonetheless, this system may be viewed as having only one mode (the impulse response function) knowledge of which simplifies immensely the system representation. This observation motivates the search for a small number of principal modes in the case of nonlinear or spike-generating systems [4], [5]. The validity of this postulate in a given application can be readily checked by the ability (or not) of a reduced model representation containing only the principal modes to explain the input–output data.

We seek to address the practical question of how to determine the minimum number of modes in the context of 2nd-order Volterra–Wiener systems (see Appendices I and II). To this purpose, we follow a two-step approach based on a kernel estimation technique that employs Laguerre expansions and estimates the expansion coefficients via least-squares fitting. The key aspects of this kernel estimation technique are summarized in Appendix I. The advantages accrued by the advent of this technique are: 1) it yields rather accurate kernel estimates from short data-records, even in the presence of significant data-contaminating noise; 2) it removes the strict requirement of white-noise inputs (indispensable in the Wiener approach previously used) so long as all significant terms are included in the model; 3) kernel representation becomes rather compact, while the computational requirements remain modest. These advantages are all critical for the successful application of the proposed modeling approach. Having obtained the Laguerre coefficient estimates of the 1st- and 2nd-order kernels, we write the system output in a quadratic form involving the (symmetric) coefficient matrix $\boldsymbol{C}$ defined in Appendix II. Eigen-decomposition of this matrix allows the selection of the most significant eigenvectors (corresponding to the most significant eigenvalues) as the Laguerre coefficients of the principal modes of the system $\{m_j(n)\}$ which are the NM's involved in the model of Fig. 1. This procedure is detailed in Appendix II.

Motivated by practical considerations, we have limited the search for possible NM's to those residing in the 1st- and 2nd-order kernels. Note, however, that this does not constrain the order of the final model of the overall system. The procedure outlined above is illustrated with computer simulated examples in Section IV.

III. Identification of the Multi-Input Threshold

Assuming that the NM's of a given system have been determined, we turn to the identification of the multi-input threshold (MT) operator of the model in Fig. 1. This is a binary operator (output 0 or 1) receiving K real-valued inputs $\{u_j(n)\}(j = 1, \cdots, K)$ which are the outputs of the K previously selected NM filters:

$$y(n) = T[u_1(n), \cdots, u_K(n)] \tag{1}$$

where $T[\cdot]$ represents the binary operator MT. In practice, (1) will only hold approximately due to the modeling errors introduced by the omission of the less significant eigenvalues of $\boldsymbol{C}$ (discussed in Appendix II). If we assume that Eq. (1) is exact, then we can define the "trigger regions" (TR) of the system as the locus of points $(u_1^*, \cdots, u_K^*)$ in the K -dimensional space that correspond to an output spike [4]. These TR's can be identified from input-output data by mapping the "trigger points" (TP) $(u_1^*, \cdots, u_K^*)$ in the K -dimensional space, as demonstrated in Section IV. When sufficient data are available, the TR's of the system will be adequately filled by TP's and, thus, the operational characteristics of the MT operator will be defined. If the expression in (1) is approximate, either because the modes are poorly estimated or their number is inadequate, then "blurring" of the boundaries of the TR's may occur. These two potential problems can be addressed by increasing the length of the data-records or the number K of employed NM's, respectively (although, in the latter case, we inadvertedly increase the dimensionality of the TR's). The tradeoff between accuracy and complexity will have to be determined by the requirements of each particular application. Clearly, retaining a small number of NM's (up to 3) facilitates the visualization of the TR's of a system. On the other hand, increasing the length of the data-records (if experimentally feasible) improves estimation accuracy and adds minimal computational burden to the identification task.

In closing this section, it is useful to draw the connection between the output expressions of (1) and the Volterra expansion discussed in Appendix I. The Volterra expansion is an analytic (power series) approximation of the non-analytic hard-threshold operator T. As discussed in previous publications [4], [5], the two representations can be connected through the application of a hard-threshold Θ upon the output of the Volterra model $y = f(\cdot)$. Then the TR's are demarcated by the equation:

$$f(v_1, v_2, \cdots v_j, \cdots) = \Theta. \tag{2}$$

Upon selection of the K principal modes $\{u_j\}$, the output of the Volterra model can be written as

$$\tilde{y}(n) = F(u_1, u_2, \cdots u_K) + \varepsilon(n) \tag{3}$$

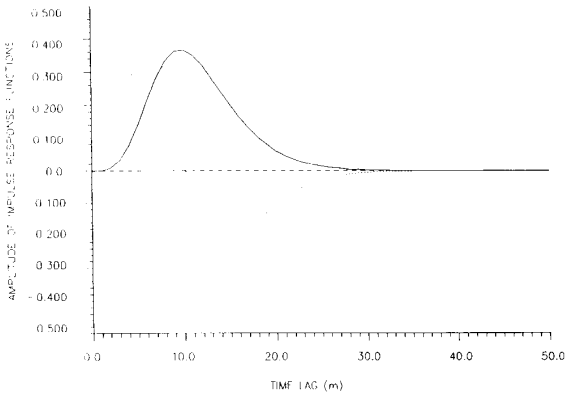


Fig. 2. The two impulse response functions used in the simulation example plotted over 50 lags: g_1 (solid line) yields a finite-bandwidth measure of stimulus intensity, and g_2 (dotted line) yields a finite-bandwidth measure of rate of change in stimulus intensity.

where $\varepsilon(n)$ represents the aforementioned approximation error which may blur the demarcation of the TR's of the model. In this case, the boundaries of the TR's may not be sharply defined as "trigger lines" but, instead, they may emerge as "trigger zones" where probability of firing is between 0 and 1.

The evaluation of the TR's was described above in a graphical/geometric manner, but it may also be done analytically by least-squares fitting of a multinomial expression of the estimated mode outputs $\{u_j(n)\}$ to the system output $y(n)$. This fitting requires simply linear regression (since the multinomial expression is linear in the unknown coefficients—see Appendices I and II) and yields an analytical expression approximating $F(\cdot)$ in (3) as a multi-nomial of specified degree. This degree ought to be specified in practice after visual inspection of the graphical estimate of the TR's discussed above. The accuracy of the obtained TR's may be readily evaluated by the predictive ability of the resulting model.

IV. Computer Simulated Examples

In order to illustrate the ideas presented in the previous sections and verify the efficacy of the proposed approach, we simulate examples of systems having the output equation:

$$y(n) = T[\alpha_1 v_1(n) + \alpha_2 v_2(n) + \beta_1 v_1^2(n) + \beta_2 v_2^2(n)] \tag{4}$$

where v_1, v_2 are the convolutions of a Gaussian white noise (GWN) input $x(n)$ with the impulse response functions $g_1(n)$, $g_2(n)$, respectively, shown in Fig. 2. Note that v_1 provides a finite-bandwidth measure of stimulus intensity and v_2 provides an estimate of the derivative of v_1. The coefficients (α_1, α_2, β_1, β_2) determine the relative importance of the two modes and the effect of the quadratic nonlinearities. For instance, a system that is more rate-sensitive (mode #2) than intensity-sensitive (mode #1) will have large α_2 relative to α_1, and if the rate-sensitivity is independent of sign (i.e., increase or decrease of the same magnitude yields the same effect) then "full-wave rectification" of the rate-sensitive mode output (v_2) can be represented by a large coefficient β_2 of the square

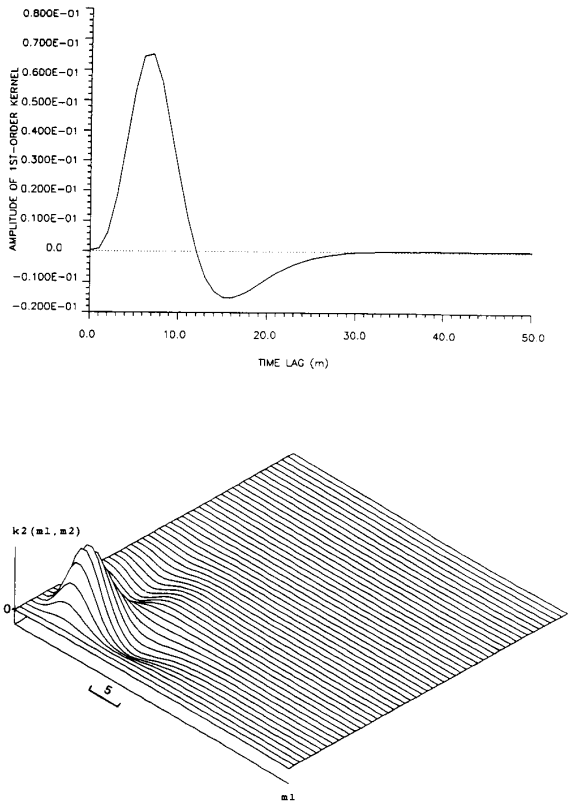


Fig. 3. The exact 1st-order kernel plotted over 50 lags (top), and the exact 2nd-order kernel plotted over 50 lags in each dimension (bottom) for the single-mode simulated system (see text). The minimum and maximum values of the 2nd-order kernel graph are -0.0027 and 0.0111, respectively.

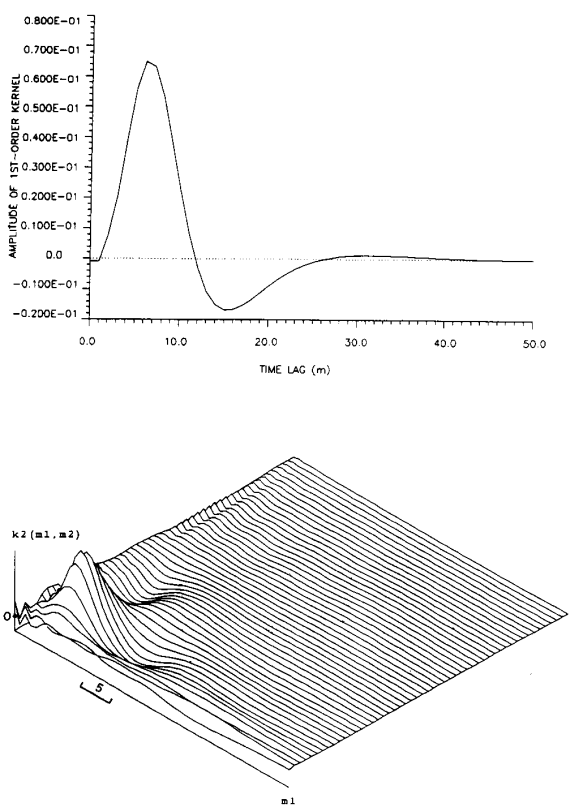


Fig. 4. The 1st-order (top) and 2nd-order (bottom) kernel estimates for the single-mode system obtained by use of the Laguerre expansion technique (LET). The minimum and maximum values of the 2nd-order kernel graph are -0.0035 and 0.0118, respectively. Comparison with their exact counterparts, shown in Fig. 3, demonstrates the efficacy of the LET in kernel estimation.

term v_2^2. This model form can be extended to multiple modes and higher degree of nonlinearities (although the proposed mode estimation procedure is based only on the 1st- and 2nd-order kernel estimates). Note that the operator $T[\cdot]$ is a hard-threshold nonlinearity that possesses a refractory period of one discrete-time point in these simulations. The GWN input data (4096 datapoints) and the resulting output spike-train are used to estimate the 1st- and 2nd-order kernels of each simulated system via the Laguerre expansion technique (LET) discussed in Appendix I. The obtained Laguerre coefficient estimates are used to determine the principal dynamic modes (neuronal modes) of each system, following the procedure outlined in Appendix II. Finally, the "trigger regions" (TR) of each system are estimated as outlined in Section III.

We begin by simulating a simple system without nonlinearities prior to the operator $T[\cdot]$, using the coefficient values: $\alpha_1 = 1$, $\alpha_2 = 2$, $\beta_1 = 0$, $\beta_2 = 0$. This represents a case of linear formation of the composite potential at the axon hillock which reduces to a single-mode system [4]. The exact 1st- and 2nd-order kernels of this system are shown in Fig. 3 (note that high order kernels arise because of the threshold nonlinearity). The obtained kernel estimates using LET (for $L = 10$ and $\alpha = 0.4$) are shown in Fig. 4 and exhibit good agreement with their counterparts in Fig. 3, demonstrating the efficacy of LET. The matrix $\boldsymbol{C}$ of the obtained Laguerre coefficient estimates (see Appendix II) is shown graphically in Fig. 5 in 3-D perspective form. Note that a number of Laguerre functions $L = 10$ is adequate in this case, even though the kernels are extending over 50 time-lags (i.e., a traditional time-domain based kernel estimation technique would have to evaluate 1326 distinct kernel values, as opposed to 55 distinct Laguerre coefficients evaluated by LET), demonstrating the potential compactness of the Laguerre expansion approach discussed previously. Following the procedure outlined in Appendix II, eigen-decomposition of this matrix $\boldsymbol{C}$ yields only one principal mode (i.e., it has only one significant eigenvalue with the next largest eigenvalue being about one order of magnitude smaller) shown in Fig. 6 (solid line) along with the 1st-order kernel estimate (dotted line). Note the proximity of the two waveforms, as anticipated by the underlying theory [5].

Since this system has a single mode, the corresponding MT operator has a single input and its TR's can be easily defined by use of scalar thresholds estimated by the data. The TR of this particular system is the range of mode output values greater than a scalar threshold $\Theta = 0.12$. The estimated mode and threshold can be used to predict the system response to any (arbitrary) stimulus by the use of the model of Fig. 1. An illustration of this is given in Fig. 7, where the actual

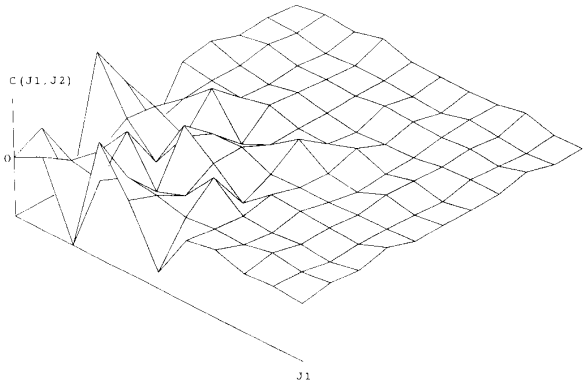


Fig. 5. The 11 × 11 symmetric matrix $\boldsymbol{C}$ of Laguerre coefficient estimates (shown in 3-D perspective form) for the single-mode system, incorporating all estimation results of the 0-th, 1st- and 2nd-order kernels. Note that $L = 10$ in this example and the values of the first row (or first column) of this symmetric matrix are the 0th order term (first point) and the Laguerre expansion coefficients of the 1st-order kernel estimate (subsequent 10 values). The remaining (10 × 10) values of the matrix correspond to the 2nd-order kernel estimate (see Appendix II). The minimum and maximum coefficient values in this graph are -0.0429 and 0.0433, respectively.

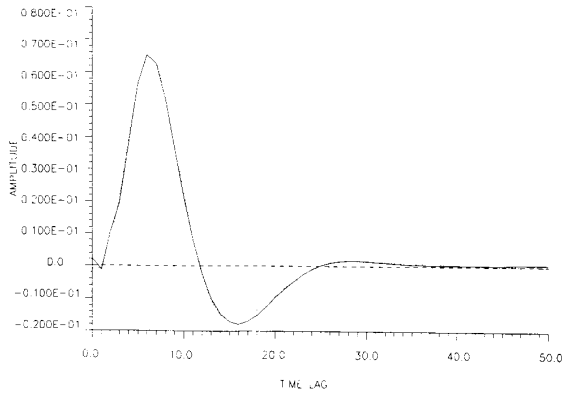


Fig. 6. The estimated "neuronal mode" for the single-mode system (solid line) and the 1st-order kernel estimate of the same system (dotted line). The remarkable agreement between the two curves confirms the theoretical expectations and demonstrates the efficacy of the proposed kernel estimation and mode estimation techniques.

response (trace 1) and the model prediction (trace 2) are shown in perfect agreement over a segment of data generated for a random stimulus (different from the one used for the estimation of the model). This remarkable predictive ability of the model corroborates the efficacy of the proposed approach.

The second example simulates a system performing additional full-wave rectification of the signals v_1 and v_2, by including the terms v_1^2 and v_2^2 in the argument of operator $T[\cdot]$, i.e., using the coefficients $\alpha_1 = 1, \alpha_2 = 1, \beta_1 = 1, \beta_2 = 1$ in (4). This represents a case of nonlinear formation of the composite potential at the axon hillock [4], which exhibits an elliptical "trigger line" (TL). Application of the proposed methodology to these data yields a new set of 1st- and 2nd-order kernel estimates, shown in Fig. 8. Comparison with the kernels of the previous system (Fig. 4) indicates marked differences, especially in the 2nd-order kernels, reflecting the different functional characteristics of the systems. Eigen-decomposition of the corresponding Laguerre coefficient matrix yields two neuronal modes (NM) shown in Fig. 9, which correspond to the two most significant eigenvalues (note that the next largest eigenvalue is about one order of magnitude smaller). As anticipated by the theory (see Appendix II), the estimated NM's, $m_1(n)$ and $m_2(n)$, can be closely approximated by linear combinations of the original impulse response functions $g_1(n)$ and $g_2(n)$ as

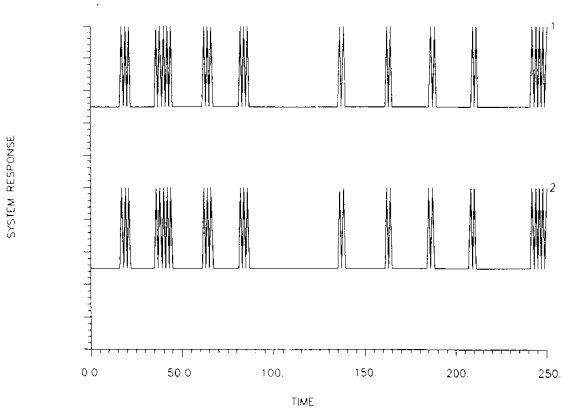


Fig. 7. The spike-train response of the single-mode system (trace 1) and the model prediction (trace 2) for a random stimulus different from the one used for estimation of the model. The excellent agreement demonstrates the predictive ability of the model.

$$m_1(n) \simeq 0.56g_1(n) + 0.78g_2(n)$$
$$m_2(n) \simeq 0.78g_1(n) - 0.69g_2(n). \quad (5)$$

The outputs $u_1(n)$ and $u_2(n)$ of the two NM's are computed as convolutions of the input $x(n)$ with $m_1(n)$ and $m_2(n)$, respectively. The "trigger points" (u_1^*, u_2^*) are mapped on the (u_1, u_2) plane, as shown in Fig. 10. The emergence of a TL is evident as the region outside the elliptical TL shown in Fig. 10, an anticipated by the theory. It was also suggested in Section III that an analytical expression for TR can be obtained by least-squares fitting of a multinomial to the output data. In this example, the multinomial is a binomial of second degree. To confirm the predictive ability of the resulting model, we show in Fig. 11 a segment of the system response along with the model prediction. The agreement is very good, with the model failing to predict only 5 out of the 46 output spikes.

An example of a system with hyperbolic TL is obtained when β_1 and β_2 have opposite signs. For instance, if $\alpha_1 = 1, \alpha_2 = 1, \beta_1 = -1, \beta_2 = 1$ in (4), then the TR' shown in Fig. 12 results, which is demarcated by a two-sided hyperbolic TL. The kernels and the NM's of this system are, of course, different from the previous examples. Thus different forms of TR's and NM's will result from systems with different nonlinearities prior to the application of the hard threshold. This offers the versatility required for modeling diverse neural systems.

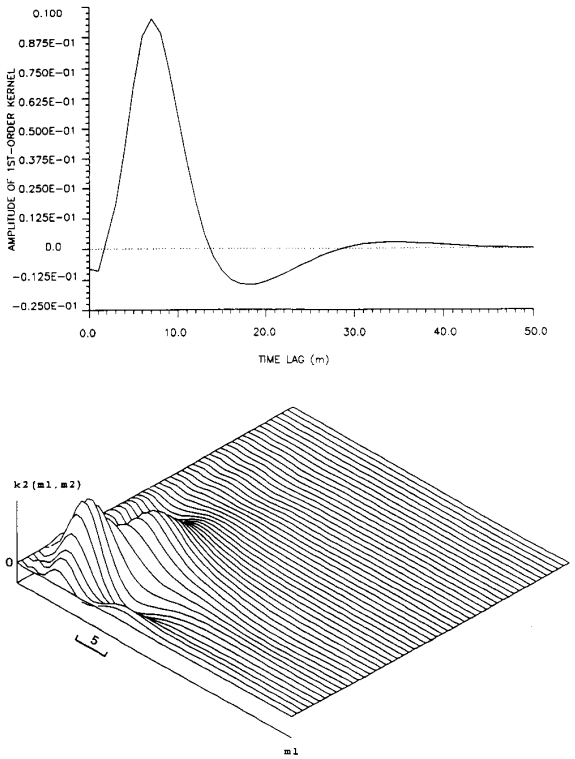


Fig. 8. The 1st-order (top) and 2nd-order (bottom) kernel estimates of the simulated two-mode system with elliptical "trigger line." Note the distinct differences in the 2nd-order kernel between this system and the single-mode system (Fig. 4).

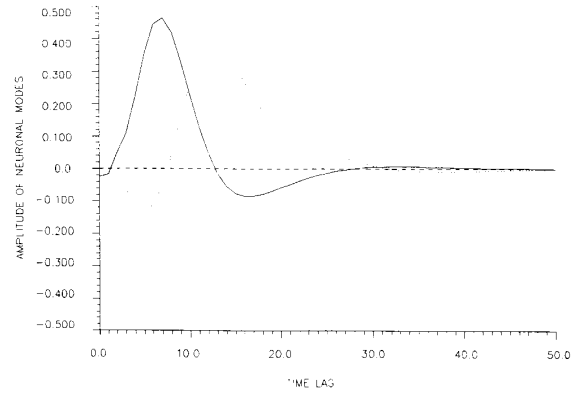


Fig. 9. The estimates of the two neuronal modes of the system with elliptical "trigger line." Note that these two modes are linear combinations of g_1 and g_2 in Fig. 2, as indicated by Eq. (5).

Let us consider an example where the nonlinearity is of degree higher than second:

$$y(n) = T[v_1 + v_2 - v_1^2 - v_2^2 + v_1^3 - v_2^3 - v_1^4 + v_2^4]. \quad (6)$$

Two significant eigenvalues are identified in the matrix of the Laguerre coefficients of this system as well. The estimated TR of this system is shown in Fig. 13, clearly demarcated by a TL that is of degree higher than second.

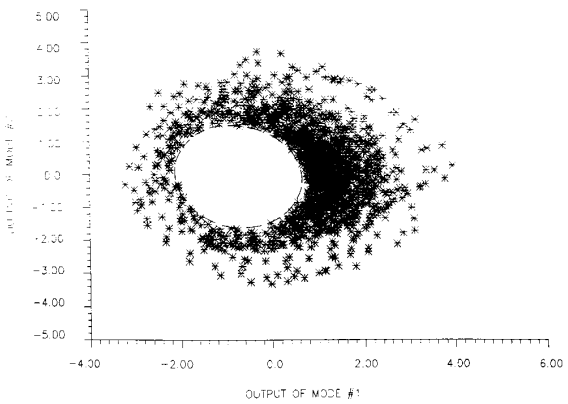


Fig. 10. Plot of the "trigger points" (asterisks) in the 2-D plane (u_1, u_2) defined by the outputs of the two modes of the system with elliptical "trigger line." The latter can be also estimated analytically by least-squares fitting (see text) and shown with dashed line. The "trigger region" is the area outside the elliptical "trigger line," as anticipated by the theory in this example.

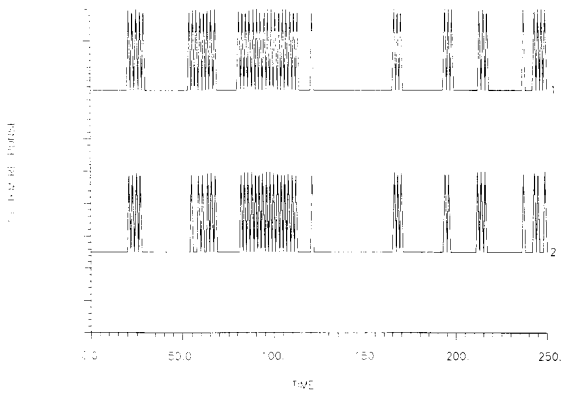


Fig. 11. The spike-train response of the two-mode system with elliptical "trigger line" (trace 1) and the two-mode model prediction (trace 2) for a random stimulus segment. The agreement is very good (with the model failing to predict only 5 out of 46 output spikes in this segment) attesting to the efficacy of the proposed modeling approach.

This example illustrates the fact that systems of arbitrary order of nonlinearity may be modeled by this approach. Note that a *predictive* model is obtained by this approach in all cases where the NM's and the TR's can be reliably estimated by the available input-output data.

Finally, the proposed method is not limited to the use of GWN inputs, as long as the employed experimental input is sufficiently "rich" in information content (i.e., broadband and sustained). In order to demonstrate this point, we repeat the simulation for $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta_1 = 1$, $\beta_2 = 1$ in (4) with a nonwhite broad-band random input whose spectrum is shown in Fig. 14. The obtained NM's are shown in Fig. 15, exhibiting good agreement with the previously estimated NM's using GWN input (Fig. 9). The resulting TR is shown in Fig. 16 and compares well with its counterpart in Fig. 10, demonstrating the efficacy of this approach when nonwhite (but broad-band) inputs are used. These computer simulated examples illustrate the use of the proposed approach for modeling spike-

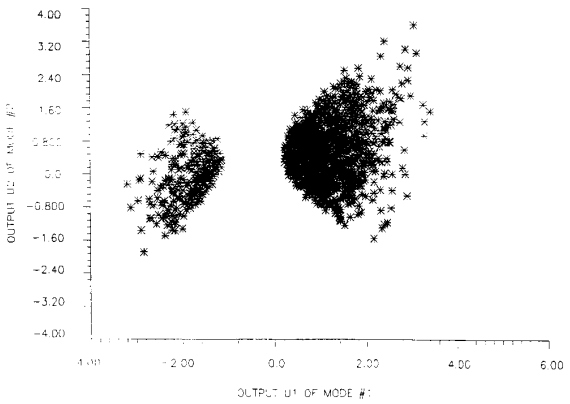


Fig. 12. Plot of the "trigger points" (asterisks) in the 2-D plane (u_1, u_2) defined by the outputs of the estimated two-modes of the simulated system with hyperbolic "trigger line". The hyperbolic "trigger region" of this example is clearly depicted, validating the estimation and modeling method.

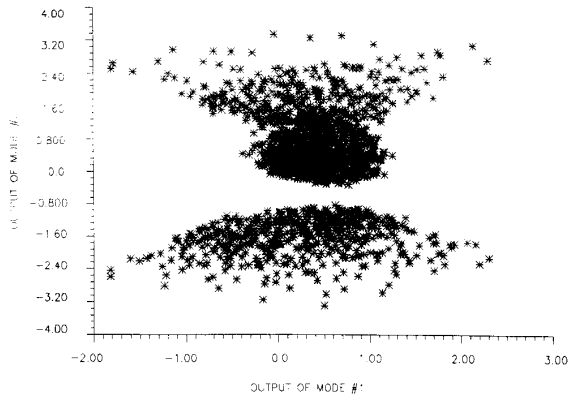


Fig. 13. Plot of the "trigger points" (asterisks) in the 2-D plane defined by the outputs of the two-modes of the simulated 4-th order system that demonstrates the applicability of this approach to higher-order systems. The 4-th degree "trigger line" is demarcated by the intricate boundary of the clusters of plotted points.

output systems. The application of this methodology to actual experimental data from neural systems is currently in progress.

V. Conclusions

A methodology for modeling a class of spike-output neural systems from input–output data is proposed, which makes use of "neuronal modes" and "multi-input threshold" operators. The method is applicable to systems with observable broadband and sustained input-output signals. The neuronal modes are determined by eigen-decomposition of the coefficient matrix obtained from the Laguerre expansions of the 1st- and 2nd-order kernels of the system. The characteristics of the multi-input threshold (binary) operator are determined by the "trigger regions" defined on the basis of the data. The method may be applied to systems of arbitrary order of nonlinearity, as long as a small number of neuronal modes is sufficient for describing the essential system dynamics. Practical estimation of the neuronal modes is possible when the data are collected

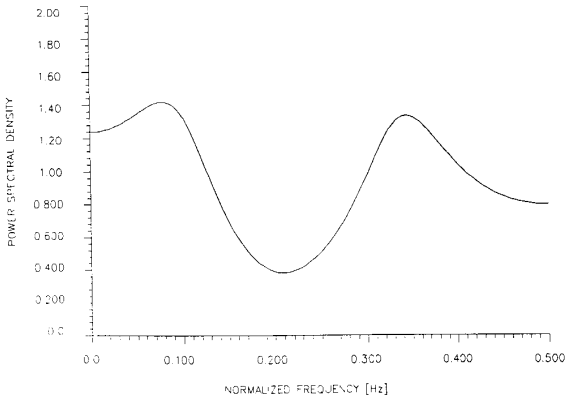


Fig. 14. The spectrum of the non-white random input used in the simulations.

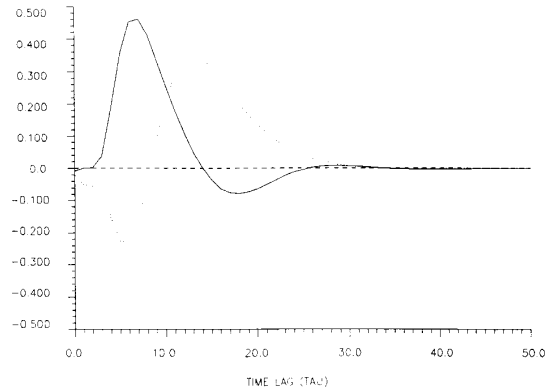


Fig. 15. The estimated neuronal modes of the two-mode system with elliptical "trigger line" when the non-white input is used. Comparison with the estimates obtained for a white input (Fig. 9) supports the applicability of this approach for nonwhite broadband inputs.

using broadband and sustained inputs (not necessarily white noise, although the latter enhances the estimation process). This approach is offered as a reasonable compromise between modeling complexity and prediction accuracy. It is hoped that it will provide a common methodological framework for modeling the broad class of neural systems which satisfy the conditions outlined in Appendix I. Note that autonomous neural systems with no discernible or observable input cannot be modeled by this approach.

Appendix I

This Appendix summarizes the basic mathematical foundation of this approach and the pivotal kernel estimation technique using Laguerre expansions. The general input–output relation of a nonlinear time-invariant dynamic system (in discrete time) is given by the Volterra series [6]:

$$y(n) = k_o + \sum_m k_1(m)x(n-m) + \sum_{m_1}\sum_{m2} k_2(m_1, m_2) \cdot x(n-m_1)x(n-m_2) + \cdots \quad (7)$$

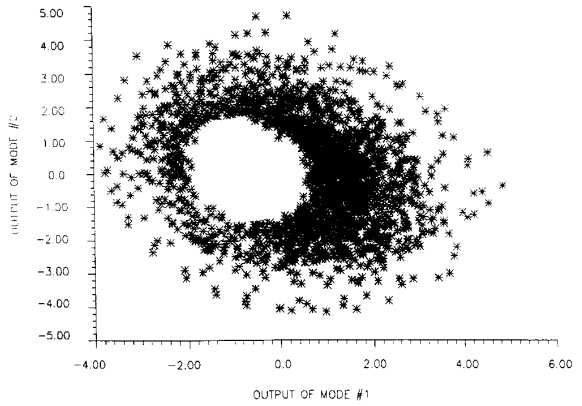


Fig. 16. The "trigger region" depicted by the estimated model with nonwhite input for the system with elliptical "trigger line." Comparison with Fig. 10 demonstrates the efficacious use of nonwhite (broadband) inputs.

where $x(n)$ is the input and $y(n)$ the output of the system. The series may extend to the appropriate (nonlinear) order for each system. The Volterra kernels $(k_o, k_1, k_2, \cdots)$ describe the dynamics of the system at each order of nonlinearity. Note that k_o represents a constant (offset) value and k_1 represents the 1st-order (linear) dynamics of the system. Kernels of higher order represent system nonlinearities (of the respective order) and they are symmetric functions (i.e., invariant under permutation of their arguments). Expansion of the Volterra kernels on the Laguerre basis $\{b_j(m)\}$ transforms (7) into the multinomial power series expression:

$$\begin{aligned} y(n) &= c_o + \sum_j c_1(j) v_j(n) \\ &\quad + \sum_{j_1} \sum_{j_2} c_2(j_1, j_2) v_{j_1}(n) v_{j_2}(n) + \cdots \\ &= f(v_1, v_2, \cdots, v_j, \cdots) \end{aligned} \tag{8}$$

where

$$v_j(n) = \sum_m b_j(m) x(n-m) \tag{9}$$

and $c_1(j), c_2(j_1, j_2), \cdots$ represent the Laguerre expansion coefficients of the 1st- and 2nd-order kernel, respectively. Note that $c_o = k_o$ and $c_2(i, j) = c_2(j, i)$. The unknown expansion coefficients can be estimated in practice by linear regression of the output data $y(n)$ on the terms of the multinomial expansion of (8), as long as it is finite. The terms of the multinomial expansion depend on the signals $v_j(n)$ given by (9) as convolutions of the input data with the selected Laguerre discrete-time functions

$$\begin{aligned} b_j(m) = {} & \alpha^{(m-j)/2} (1-\alpha)^{1/2} \sum_{k=0}^{j} (-1)^k \\ & \cdot \binom{m}{k} \binom{j}{k} \alpha^{j-k} (1-\alpha)^k \qquad (m \geq 0) \end{aligned} \tag{10}$$

where α is the discrete-time Laguerre parameter $(0 < \alpha < 1)$ which determines the rate of exponential asymptotic decline of these functions. The key variables $\{v_j(n)\}$ can be computed more efficiently by use of the recursive relation:

$$v_j(n) = \sqrt{\alpha}\, v_j(n-1) + \sqrt{\alpha}\, v_{j-1}(n) - v_{j-1}(n-1) \tag{11}$$

initialized by

$$v_o(n) = \sqrt{\alpha}\, v_o(n-1) + \sqrt{1-\alpha}\, x(n). \tag{12}$$

These computations can be performed rather fast, for $n = 1, \cdots, N$ and $j = 0, 1, \cdots, L-1$; where L is the total number of Laguerre functions used in the kernel expansion. The choice of the Laguerre parameter α is rather critical in achieving efficient kernel expansions (and, consequently, fast and accurate kernel estimation) and its judicious selection is based on the effective memory (size) of the system kernels and L. The efficacy of this technique is demonstrated with computer simulations in Section IV.

If "L" Laguerre functions are selected, then for a Q-th order system the total number of unknown Laguerre coefficients that need to be estimated is $(Q+L)!/(Q!L!)$, when the kernel symmetries are taken into account. Due to the difficulty of representation (multidimensional) of high-order kernels, kernels only up to 3rd-order have been estimated thus far. Clearly, the computational burden increases rapidly with L and, therefore, this approach is more attractive when the kernels can be represented by relatively few Laguerre functions. The determination of the appropriate L for a given system is based on successive trials with increasing L and evaluation of the resulting reduction in mean-square error of model prediction (i.e., the traditional approach to regression model order determination). Extensive computer-simulated examples have shown that this kernel estimation technique offers significant advantages over previous ones, which are summarized in the main text.

The Wiener approach to this problem entails the use of a Gaussian white-noise (GWN) input $x(n)$, which tests the system over all frequencies and amplitudes of interest. Furthermore, Wiener orthogonalized the Volterra series (7) when the input is GWN and suggested ways by which the unknown kernels of the system can be estimated from input-output data [6], [7] using co-variance estimates (the simplest implementation involves high-order cross-correlations). The resulting set of Wiener kernels is, in general, distinct from the Volterra kernels of a given system. A great deal of work has been performed and published in the last 20 years on the theory and application of the Wiener approach to physiological system modeling (see, for instance, [3], [6]). Relevant to the objectives of this paper, the following are the main Wiener contributions: 1) the proposition that a GWN input will test a nonlinear system exhaustively over the entire frequency and amplitude range, and thus extract all information necessary to identify (model) the system; 2) the suggestion that a nonlinear system of the Volterra–Wiener class can be modeled by a linear filter-bank (with impulse response functions forming a complete basis) feeding into a multi-input static nonlinearity as shown in Fig. 17. The latter suggestion is corroborated by (8) and (9), where $f(\cdot)$ represents the multi-input static nonlinearity MN and $\{b_j\}$ represent the impulse response functions of the linear filters $\{L_j\}$, in Fig. 17.

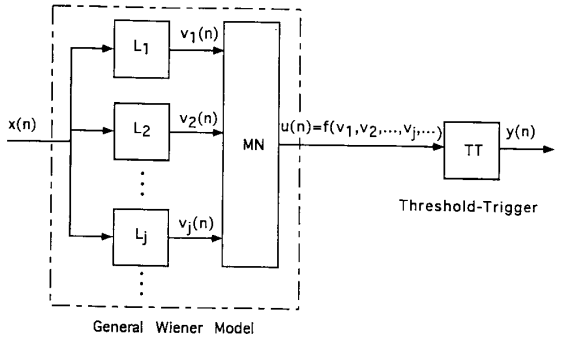


Fig. 17. Block-structured model of the general Wiener system followed by a threshold-trigger operator (TT) that generates the output spikes. The filter-bank (L_1, L_2, $\cdots$) forms a complete basis of the system kernels. The multi-input static nonlinearity MN must be representable (or approximated) by a multivariable power series.

Thus this modeling approach is applicable to all single-valued systems with square-integrable (or square-summable in discrete time) kernels and nonlinearities which can be represented or approximated by multinomial power series expansions. All systems satisfying these weak conditions belong to the Volterra–Wiener class and can be decomposed in the model form of Fig. 17. This decomposition was adapted to spike-output systems [5] by appending a theshold-trigger operator TT to the static nonlinearity MN as depicted in Fig. 17. Note that the operator TT includes a refractory period. It is evident that the class of systems representable by this model form is extremely broad. In practical terms, however, this model form is useful only when it is possible to reduce the number of filters in the bank to a few "principal dynamic modes," which are the "neuronal modes" (NM) proposed in this paper. Subsequently, the combined nonlinearities MN and TT are replaced by the proposed multi-input threshold (MT) operator receiving inputs from the outputs of the selected neuronal modes NM shown in Fig. 1 (see main text).

APPENDIX II

This Appendix outlines the procedure by which the principal dynamic modes (neuronal modes) are determined from 1st- and 2nd-order kernel estimates. Consider the Laguerre expansion coefficients estimated for a 2nd-order Volterra model, as described in Appendix I. We form the following (symmetric) coefficient-matrix $\boldsymbol{C}$ using the estimated coefficients: $\{c_1\}$ and $\{c_2\}$ of the Laguerre expansions of the 1st- and 2nd-order kernels, respectively (c_o is the zero-order constant term):

$$\boldsymbol{C}\begin{bmatrix} c_o & c_1(1)/2 & c_1(2)/2 & \cdots & c_1(L)/2 \\ c_1(1)/2 & c_2(1,1) & c_2(1,2) & \cdots & c_2(1,L) \\ c_1(2)/2 & c_2(2,1) & c_2(2,2) & \cdots & c_2(2,L) \\ \vdots & \vdots & \vdots & & \vdots \\ c_1(L)/2 & c_2(L,1) & c_2(L,2) & \cdots & c_2(L,L) \end{bmatrix}. \quad (13)$$

Note that for a 2nd-order system, the output $y(n)$ can be written as (superscript "T" denotes "transpose"):

$$y(n) = \underline{v}^T(n)\boldsymbol{C}\underline{v}(n) \quad (14)$$

where the vector:

$$\underline{v}^T(n) = [1 \quad v_1(n)v_2(n) \quad \cdots \quad v_L(n)] \quad (15)$$

represents the outputs of the Laguerre filters at each time n (augmented by a constant element). Consider the eigen-decomposition of the symmetric matrix $\boldsymbol{C}$ using the matrix M of its eigenvectors:

$$\boldsymbol{C} = M^T\Lambda M \quad (16)$$

where Λ is the diagonal matrix of its (distinct) eigenvalues $\{\lambda_j\}(j = 0, 1, \cdots, L)$. Then

$$\begin{aligned} y(n) &= [M\underline{v}(n)]^T\Lambda[M\underline{v}(n)] \\ &= \underline{u}^T(n)\Lambda\underline{u}(n) \\ &= \lambda_0 u_0^2(n) + \lambda_1 u_1^2(n) + \cdots + \lambda_L u_L^2(n) \end{aligned} \quad (17)$$

where the elements $\{u_j(n)\}$ of the transformed vector $\underline{u}(n)$ are linear combinations of the elements of the vector $\underline{v}(n)$, i.e.,

$$u_j = \mu_{j,0} + \mu_{j,1}v_1(n) + \cdots + \mu_{j,L}v_L(n) \quad (18)$$

where $[\mu_{j,0}\,\mu_{j,1}\cdots\mu_{j,L}]$ is the j-th eigenvector $\underline{\mu}_j$ of $\boldsymbol{C}$. Equation (17) indicates that the relative importance of $u_j(n)$ for the system output $y(n)$ is determined by the relative magnitude of the corresponding eigenvalue λ_j. Thus inspection of the relative magnitude (absolute value) of the eigenvalues of $\boldsymbol{C}$ allows quantitative selection of the "principal modes" of the quadratic form in (17). In essence, by selecting only those eigenvectors of matrix $\boldsymbol{C}$ which correspond to eigenvalues of significant magnitude, we concentrate on those linear combinations of the elements of vector $\underline{v}(n)$ which contribute most to the output of $y(n)$. Note that, if the eigenvalues of matrix $\boldsymbol{C}$ are not all distinct, then cross-terms $\{u_iu_j\}$ will appear in (17). In practice, the selection of the significant eigenvalues/eigenvectors is based on criteria deemed suitable for each application by the individual investigator(s). Naturally, signal-to-noise ratio considerations and tradeoffs between accuracy and complexity must be taken into account in exercising this judgment.

Having selected the $K(K \leq L)$ "principal components" $\{u_j(n)\}$, we can reconstruct the K time-domain functions $\{m_j(n)\}$ whose Laguerre expansion coefficients correspond to the selected K eigenvectors $\{\underline{\mu}_j\}$:

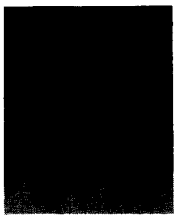
$$m_j(n) = \mu_{j,0} + \mu_{j,1}b_1(n) + \cdots + \mu_{j,L}b_L(n). \quad (19)$$

These functions $\{m_j(n)\}$ are the selected "neuronal modes" (NM) of the system, i.e. they are the impulse response functions of the minimum number K of filters in Fig. 1.

REFERENCES

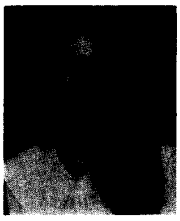
[1] T. H. Bullock, R. Orkand and A. Grinell, *Introduction to Nervous Systems*. San Francisco, CA: Freeman, 1977.
[2] C. Koch and I. Seger, Eds., *Methods in Neuronal Modeling*. Cambridge, MA: MIT Press, 1989.
[3] V. Z. Marmarelis, Ed., *Advanced Methods of Physiological System Modeling: Vol II*. New York: Plenum, 1989.
[4] V. Z. Marmarelis, "Signal transformation and coding in neural systems," *IEEE Trans. Biomed. Eng.*, vol. 36, pp. 15–24, 1989.

[5] V. Z. Marmarelis, M. C. Citron, and C. P. Vivo, "Minimum-order Wiener modeling of spike-output systems," *Biol. Cybern.*, vol. 54, pp. 115–123, 1986.
[6] P. Z. Marmarelis and V. Z. Marmarelis, *Analysis of Physiological Systems: The White Noise Approach.* New York: Plenum, 1978. (Russian translation: Mir Press, Moscow, 1981).
[7] N. Wiener, *Nonlinear Problems in Random Theory.* New York: The Technology Press of M.I.T. and Wiley, 1958.



Vasilis Z. Marmarelis was born in Mytilini, Greece, on November 16, 1949. He received the Diploma in electrical and mechanical engineering from the National Technical University of Athens in 1972 and the M.S. and Ph.D. degrees in engineering science (information science and bioinformation systems) from the California Institute of Technology in 1973 and 1976, respectively.

After two years of post-doctoral work at the California Institute of Technology, he joined the faculty of Biomedical and Electrical Engineering at the University of Southern California, where he is currently Professor and Chairman of Biomedical Engineering. He has been Director of the Biomedical Simulations Resource, a research center funded by the National Institutes of Health since 1985, which is dedicated to modeling and simulations studies of biomedical systems. His main interests are in the areas of nonlinear and nonstationary system identification and modeling, with applications to biology, medicine, and engineering systems. Other interests include spatio-temporal and nonlinear/nonstationary signal processing, and analysis of neuronal networks. He is co-author of the book *Analysis of Physiological Systems: The White-Noise Approach* (New York: Plenum, 1987; Russian translation: Mir Press, Moscow, 1981) and editor of two volumes on *Advanced Methods of Physiological System Modeling* (1987, 1989). He has published more than 60 papers and book chapters in the area of system and signal analysis.



Melissa E. Orme was born in Los Angeles, CA, on March 12, 1961. She received the B.S., M.S., and Ph.D. degrees in aerospace engineering from the University of Southern California in 1984, 1985, and 1989, respectively.

She served as a Research Assistant Professor in the Department of Aerospace Engineering at the University of Southern California from 1989 to 1993. She is currently an Assistant Professor in the Department of Mechanical and Aerospace Engineering at the University of California at Irvine. Her main interests are in the areas of dynamic modeling of mechanical systems and capillary fluid dynamics with applications to combustion and advanced materials synthesis. She has published more than a dozen papers and has been granted three patents.

Dr. Orme is a recipient of the American Association of University Women Award, and the Zonta International Amelia Earhart award.