Self-Organized Segmentation of Time Series: Separating Growth Hormone Secretion in Acromegaly from Normal Controls

Klaus Prank,* Mirko Kloppstech,* Steven J. Nowlan,[‡§] Terrence J. Sejnowski,[§¶] and Georg Brabant*

*Abteilung Klinische Endokrinologie, Medizinische Hochschule Hannover, D-30623 Hannover, Germany; [‡]Lexicus, Motorola, Palo Alto, California 94306 USA; [§]Howard Hughes Medical Institute and Computational Neurobiology Laboratory, The Salk Institute, San Diego, California 92186-5800 USA; and [¶]Department of Biology, University of California, San Diego, California 92093 USA

ABSTRACT The pulsatile pattern of growth hormone (GH) secretion was assessed by sampling blood every 10 min over 24 h in healthy subjects (n = 10) under normal food intake and under fasting conditions (n = 6) and in patients with a GH-producing tumor (acromegaly, n = 6), before and after treatment with the somatostatin analog octreotide. Using autocorrelation, we found no consistent separation in the temporal dynamics of GH secretion in healthy controls and acromegalic patients. Time series prediction based on a single neural network has recently been demonstrated to separate the secretory dynamics of parathyroid hormone in healthy controls from osteoporotic patients. To better distinguish the differences in GH dynamics in healthy subjects and patients, we tested time series predictions based on a single neural network and a more refined system of multiple neural networks acting in parallel (adaptive mixtures of local experts). Both approaches significantly separated GH dynamics under the various conditions. By performing a self-organized segmentation of the alternating phases of secretory bursts and quiescence of GH, we significantly improved the performance of the multiple network system over that of the single network. It thus may represent a potential tool for characterizing alterations of the dynamic regulation associated with diseased states.

INTRODUCTION

Pulsatile hormone secretion has been demonstrated in a large number of different endocrine systems with a pulse frequency ranging from approximately 6 to 140 pulses in 24 h (Brabant et al., 1992). This dynamic mode of secretion has important effects on target cell regulation. In perifused rat hepatocytes, glucose production is closely dependent on the frequency of stimulating glucagon pulses (Weigle and Goodner, 1986). Moreover, it could be demonstrated in single hepatocytes that information encoded in the temporal pattern of an extracellular (adrenergic) stimulus is transferred across the plasma membrane and translated into distinct intracellular information encoded in the frequency and amplitude of Ca^{2+} transients (Schöfl et al., 1993). Other well-characterized examples are the frequency-encoded regulation of pituitary hormones by hypothalamic releasing factors. Thyrotropin (TSH) secretion depends on the frequency of thyrotropin-releasing hormone (TRH) stimulation, which differentially regulates transcription of the α- and β-subunits of TSH (Haisenleder et al., 1992). A comparable frequency-dependent subunit control is observed for luteinizing hormone (LH) and follicle-stimulating hormone (FSH) by gonadotropin-releasing hormone (GnRH) (Knobil, 1980; Shupnik, 1990). This has important consequences for the treatment of patients with hypothalamic failure of GnRH secretion, such as in Kallman's syndrome, where only an adaptation of the physiological rhythm of pulsatile GnRH stimulation is able to therapeutically correct the disorder (Santoro et al., 1986). Regulation of cellular functions by the frequency of hormone pulses could also be demonstrated on the molecular level for the growth hormone (GH)-dependent expression of mRNA for insulin-like growth factor I (IGF-I) and hepatic steroid hydroxylases in the rat. In hypophysectomized rats, IGF-I secretion and mRNA expression are stimulated significantly better by dividing the same cumulative dose of GH per day into two to six distinct pulses than by continuous application. This leads to a specific pattern of hepatic enzyme expression, a significantly larger weight gain, and somatic growth (Clark et al., 1985; Isgaard et al., 1988; Waxman et al., 1991).

Time series prediction has been used to capture regularities in the temporal pattern of complex time series, particularly in the context of nonlinear dynamical systems, and to separate deterministic from random behavior (Sugihara and May, 1990; Tsonis and Elsner, 1992; Weigend and Gershenfeld, 1993). This technique tries to predict the future dynamics of a time series from a number of past values. Differences in the temporal dynamics of a system are then reflected in an altered predictability. Such predictive approaches have been effectively used in several biological systems (Blinowska and Malinowksi, 1991; Chang et al., 1994; Lefebvre et al., 1993; Schiff et al., 1994; Sugihara, 1994; Scott and Schiff, 1995) and may be applied for short time series containing a very limited number of data points (Sugihara and May, 1990; Tsonis and Elsner, 1992). In the Santa Fe Institute competition on time series prediction, the best predictions for a number of data sets from various fields were achieved by neural network approaches

Received for publication 11 December 1995 and in final form 7 March 1996.

Address reprint requests to Dr. Klaus Prank, Abteilung Klinische Endokrinologie, Medizinische Hochschule Hannover, D-30623 Hannover, Germany. Tel.: 49-511-532-3827; Fax: 49-511-532-3825; E-mail: ndxdpran@rrzn-user.uni-hannover.de.

0006-3495/96/06/2540/08 $2.00

(Weigend and Gershenfeld, 1993). These approaches have the advantage of broad linear and nonlinear approximation abilities, without making specific assumptions about the system under study, in contrast to explicit models using a set of differential equations. We recently applied neural networks to endocrine time series prediction to clearly separate the secretory dynamics of parathyroid hormone (PTH) in a group of osteoporotic patients from that in healthy controls, demonstrating a significantly higher predictability of the PTH secretory dynamics in healthy subjects than in the osteoporotic group (Prank et al., 1995). In 24-h profiles of PTH serum concentrations, we could additionally demonstrate a switching behavior between two dynamic phases, a phase of high and one of low predictability. Based on these results, a single neural network predictor might not have been optimal for the prediction of time series with different switching dynamics.

Divide-and-conquer strategies are effective methods for solving complex problems, dividing them into simpler problems, the solutions of which can be combined to yield a solution to the complex problem (Jacobs et al., 1991; Jordan and Jacobs, 1994; Nowlan, 1990; Pawelzik et al., 1996). The Mixture of Experts architecture (Jacobs et al., 1991; Nowlan, 1990) was proposed as an approach to learning a task decomposition in parallel in a neural network context. Here we focus on the adaptive capabilities of mixtures of local experts (Jacobs et al., 1991), for the prediction of 24-h GH time series in patients with a GH-producing pituitary tumor (acromegaly) and normal controls. The network system was trained to perform time series prediction of pooled data from 24-h GH rhythms from a reference group of healthy controls as well as from each subgroup. Its predictive abilities were then tested on an additional group of healthy controls as well as on acromegalic patients. To avoid possible differences in the analytical results simply caused by the higher mean GH pulse amplitude in acromegalic patients, we also compared the physiologically enhanced GH secretion in fasting normals with the pharmacologically reduced secretion in acromegalic patients treated with a somatostatin analog (octreotide). The two groups have comparable mean GH serum levels. Using this approach, we found that the neural network system had performed a self-organized segmentation of the secretory dynamics by learning time series prediction without any prior knowledge of the form of a pulse or the model of secretion. It significantly separated the GH secretory dynamics in acromegaly under basal conditions and under octreotide treatment from that in normal controls in basal and fasting states.

MATERIALS AND METHODS

Subjects and study design

Ten young, lean male subjects (aged 22–33 years, body mass index (BMI) 23.8 kg/m^2) and six patients (three males and three females) with clinical and biochemical diagnosis of acromegaly (aged 27–54 years) not cured by previous surgical and/or radiation therapies took part in this study. The studies were approved by the local Committee on Medical Ethics, and all subjects and patients gave their informed written consent. All healthy volunteers were carefully examined to exclude underlying diseases. The healthy volunteers and the acromegalic patients were admitted to our endocrine unit at noon, and a central venous catheter was placed. Blood samples were taken every 10 min over 24 h, starting at 1800 h. Sleep was allowed between 2300 and 0600 h. Five healthy controls were studied twice, under basal metabolic conditions (Fig. 1 *a*) and after 3 days of fasting (Fig. 1 *b*) to enhance GH secretion (Ho et al., 1988), with an interval of 3–6 weeks between studies. Fasting was controlled by daily measurement of ketone bodies in the urine. The 24-h GH secretory profile in the acromegalic group was studied under two different conditions: under basal condition (Fig. 1 *c*) without any medication, followed 4 weeks later under continuous s.c. pump infusion of 300 μg long-acting somatostatin analog SMS 201-995 (SMS)/d (Fig. 1 *d*). On the fourth day of infusion, blood sampling was started over 24 h. The study design has been described in more detail previously, and some of the data of these studies have been included in the present data set (Riedel et al., 1992, 1995).

Assays

The hormone concentrations of all serum samples were measured in duplicate. Commercially available immunoradiometric assays with a sensitivity of 0.2 mU/liter were used. Serum samples below the detection limit were set to 0.01 mU/liter. Intra- and interassay coeffficient of variation ranged from 2.5% and 7.1% for low hormone concentration values to 6.0% and 8.0% for high values. The assay characteristics have been described in more detail previously (Riedel et al., 1992, 1995).

Linear separability

To test for basic linear separability of the GH time series we computed the autocorrelation function (acf) and recorded the decay time, defined as the time lag L at which the acf becomes zero. The decay time was statistically

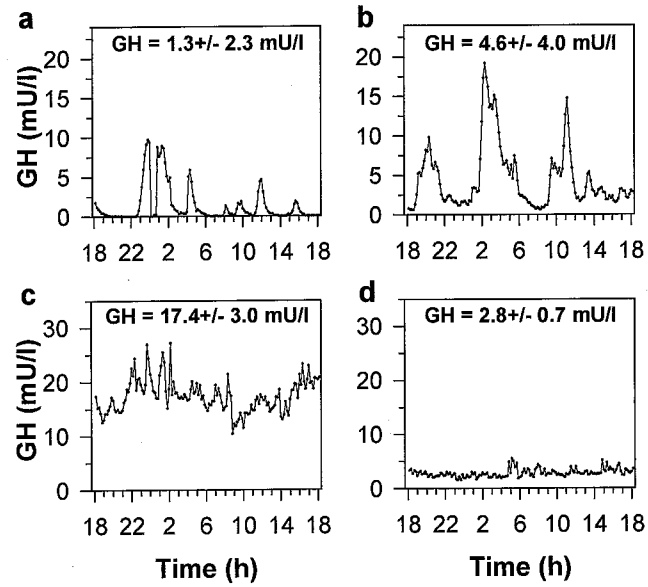

FIGURE 1 Representative 24-h GH serum concentration time series (10-min sampling interval) with attached mean 24-h GH serum concentration ± SD. Healthy control: (*a*) basal condition and (*b*) after 3 days of fasting; acromegalic patient: (*c*) basal condition and (*d*) octreotide treatment.

compared between healthy controls and acromegalic patients under either study condition. The autocorrelation function is defined as

$$\Psi(L) = \frac{\sum_{i=1}^{n-L} (x(t_i) - \langle x \rangle)(x(t_{i+L}) - \langle x \rangle)}{\sum_{i=1}^{n-L} (x(t_i) - \langle x \rangle)^2}, \quad (1)$$

where L is the time lag and the angle brackets denote the mean value of x.

Time series prediction

The next value $x(t_i)$ of a time series is predicted from m previous values using an appropriate mathematical model f:

$$x(t_i) = f(x(t_{i-1}), x(t_{i-2}), \ldots, x(t_{i-m})) + \epsilon_i, \quad (2)$$

where $x(t_{i-1})$ is the preceeding value, and ϵ_i corresponds to noise or fitting error. In our study we used feedforward neural networks, where the model f represents network architectures with linear (equivalent to autoregressive (AR) models $[f(x(t_{i-1}),\ldots, x(t_{i-m})) = \Sigma\, \alpha_k x(t_{i-k})]$) or nonlinear (sigmoidal) activation functions. By using such a model, overfitting can be controlled by applying regularization functions and cross-valdidation. These network approaches have been shown previously to be effective predictors for other short, noisy time series (Lapedes and Farber, 1987; Nowlan and Hinton, 1992; Weigend et al., 1990; Weigend and Gershenfeld, 1993).

The network models were trained (fitting the model f to the data by minimizing some error of misfit) to predict one time step (10 min) ahead using data from the past. The value predicted one time step into the future,

$$\hat{x}(t_i) = f(x(t_{i-1}), x(t_{i-2}), \ldots, x(t_{i-m})) + \epsilon_i, \quad (3)$$

was iterated back to the input of the network to predict two steps ahead:

$$\hat{x}(t_{i+1}) = f(\hat{x}(t_i), x(t_{i-1}), \ldots, x(t_{i-m+1})) + \epsilon_i, \quad (4)$$

This procedure can be repeated to predict any given number of time steps into the future. To decorrelate the prediction error estimate from the variance of the respective time series, we used the average relative variance (arv)

$$\text{arv} = \frac{\langle (x_i - \hat{x}_i)^2 \rangle}{\sigma^2(x_i)}, \quad (5)$$

where the angle brackets denote the mean over all predictions and $\sigma^2(x_i)$ denotes the variance of the measured variable. This measure was used to evaluate the predictive performance of one- and multiple-step-ahead predictions during training and testing.

Mixtures of experts

Modular, multiple network architectures (mixtures of experts), as shown in Fig. 2, have been suggested for tasks in which the set of training cases may be naturally divided into subsets that correspond to distinct subtasks. Each "expert" network learns to solve a different subtask, and the "gating" network learns to decide which expert to select for a particular case. If the appropriate division into subtasks is known in advance, it is possible to train each of the "expert" networks separately on the appropriate subtask and then train the gating network that decides which of the experts should be used for each training case (Hampshire and Waibel, 1989). Because the appropriate division into subtasks in this study is not known before analysis, we are interested in the more difficult case in which the appropriate decomposition must be discovered by the system during the training process.

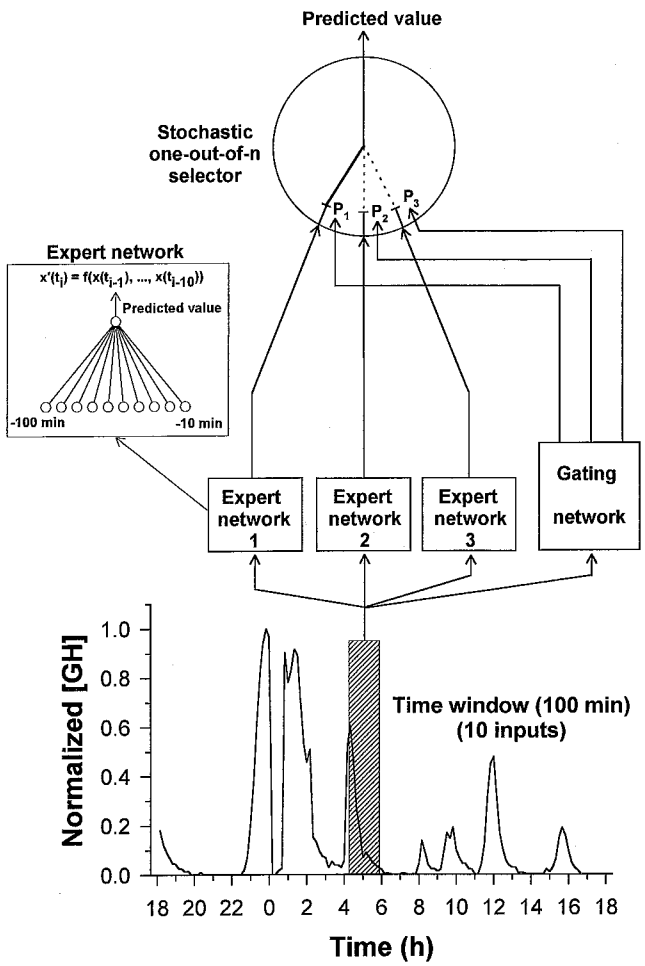

FIGURE 2 Schematic architecture of adaptive mixtures of local experts used for time series prediction. Each local expert is a feedforward neural network that has all the same architecture (10 input units, one output unit, and linear activation functions) and receive the same input. The gating network is also feedforward and receives the same input as the expert networks. It has normalized outputs $p_j = \exp(x_j)/\Sigma_i \exp(x_i)$, where x_j is the total weighted input received by output unit j of the gating network. The selector acts like a multiple input, single output stochastic switch. The probability that the switch will select the output from expert j is P_j, which is the jth output of the gating network. In this example the input pattern consists of ten normalized ([0, 1]) values of a GH serum concentration time series (100-min input window), which are used to predict the next value. These local experts are trained in parallel to perform time series prediction one time step ahead, using pooled data from a reference group (n = 5) of healthy controls. During training the experts organize themselves in such a way that each expert specializes to best predict distinct parts of the time series. The trained network system is then tested on additional untrained GH time series from healthy controls (n = 5) and acromegalic patients (n = 6). During testing the weighted average of the expert outputs (with weights assigned by the gating network) is used as the output of the mixture of experts.

The basic idea of such a system is that the gating network allocates a new case to one or a few experts, and if the output is incorrect, the weight changes are localized to these experts as well as the gating network. During the training procedure the experts compete to generate the desired output for each input pattern and adapt to a particular input pattern in proportion to their performance relative to the other expert networks. Instead of linearly combining the outputs of separate expert networks, the gating network makes a stochastic decision about which single expert to use on

each occasion (Fig. 2). As a result, the goal of a local expert network on a particular training case is not directly affected by the weights within other local experts. There is still some indirect coupling, because if some other expert changes its weights, it may cause the gating network to alter the responsibilities that are assigned to the experts. When an expert has less error than the weighted average of the errors of all experts, its responsibility for that case is increased, and when it does worse, its responsibility is decreased. If both the gating network and the local experts are trained by gradient descent, the modular network system tends to devote a single expert to each training case. Simulations on a complex vowel classification task have shown that adaptive mixtures of local experts are able to effectively decompose a problem to yield higher classification performance than a single network (Nowlan, 1991). The mixtures of experts also generalizes better from limited amounts of training data.

Training and testing of single predictive neural networks

Before training, each GH serum concentration time series was normalized to the interval [0,1] to exclude any bias caused by different mean GH pulse amplitudes in the acromegalic and control group and to account for nonlinearities in the sigmoidal activation functions. First, we evaluated the predictive performance of a large variety of single feedforward neural networks with linear as well as nonlinear (sigmoidal) activation functions. The network size ranged from 6 to 24 input units, corresponding to a time window between 1 and 4 h, 0 to 12 hidden units, and one output unit. In our computer experiments we used a regularization technique (Nowlan and Hinton, 1992) to control overfitting during training (Prank et al., 1995). The weights were updated by a conjugate gradient descent method. For linear activation functions, this method is formally equivalent to a least-squares linear fit. The different network architectures were trained to predict the next GH concentration value of a time series from its past m values, where m is the number of input units, using pooled GH time series data from the control and acromegalic group under either condition (basal and fasting or SMS treatment) and applying a "leave one out technique," in which the time series being tested for prediction is left out of the training procedure. The predictive performance of the architectures was evaluated on validation time series that had not been used for training previously. The performance of the best predicting single neural network architecture (Table 2) within each subgroup was then compared to the corresponding mixture of experts neural network system (Table 3). The simulations of single predictive neural networks were performed on a Sun SPARCstation 20 using customized C-code.

Training and testing of adaptive mixtures of local experts

The one-step ahead prediction error (arv) was evaluated in a variety of experiments for each group, using adaptive mixtures of local experts. In our simulations we varied the number of local experts between two and six and the input window size between 7 and 16 data values. Each expert consisted of a feedforward linear network without hidden units. We varied the size of the gating network from 7 to 16 input units, 0 to 5 hidden units, and we used linear as well as nonlinear (sigmoidal) activation functions. The best predictive mixture of experts system was selected from the 120 different simulations performed for each group based on the one-step prediction testing error (Fig. 3). The respective system was then used to perform some of the further analysis. The performance of the best predictive mixture of experts system within each group was statistically compared between the control and the acromegalic groups under each study condition (Table 4). Additionally, we trained the best predictive mixture of experts system using pooled reference data from healthy controls ($n = 5$). The system used for this simulations consisted of five linear experts with nine input units each and a gating network with three hidden units and sigmoidal activation functions. It was tested for its predictive ability on the remaining control and acromegalic GH time series (Fig. 4). The number of the respective expert selected to predict a given value of the time series was recorded versus time (Fig. 5). Finally, a frequency distribution of the expert selection was computed for each subgroup (Table 5). Because we do not know the appropriate division into segments of the time series (training respectively testing cases) before analysis of the GH rhythms, distinct local experts were automatically assigned to the corresponding training and testing cases by the gating network. During training the weights of the local experts as well as those of the gating network were adjusted by a conjugate gradient descent method, resulting in the final "automatic" specialization of each expert. Adaptive mixtures of local experts were simulated with customized C-code on a Sun SPARCstation 20.

TABLE 1 Decay time of autocorrelation function

Group 1 vs. group 2	Decay time (min)	p
Acromegaly, basal	252 ± 80	
Control, basal	120 ± 43	0.009
Acromegaly, basal	252 ± 80	
Control, fasting	250 ± 134	NS
Acromegaly, SMS vs.	267 ± 150	
Control, basal	120 ± 43	NS
Acromegaly, SMS vs.	267 ± 150	
Control, fasting	250 ± 134	NS

TABLE 2 Architecture of the best performing single neural network predictor

Group	Input units	Hidden units	Output units	Activation function
Acromegaly, basal ($n = 6$)	5	8	1	Sigmoidal
Acromegaly, SMS ($n = 6$)	7	5	1	Sigmoidal
Control, basal ($n = 10$)	7	10	1	Sigmoidal
Control, fasting ($n = 5$)	7	0	1	Sigmoidal

Statistics

Data are presented as mean ± SD. Statistical analysis was performed using a nonparametric Mann-Whitney-Wilcoxon rank test. A value of $p < 0.05$ was assumed as the significance level.

RESULTS

The 24-h GH profiles appear to visually differ between healthy controls and acromegalic patients under either study condition (Fig. 1). Although the healthy control is easily separable from the acromegalic patient by a lower mean 24-h GH serum concentration under basal conditions, this is not the case for the comparison between healthy subjects

TABLE 3 Single vs. mixture of experts predictor

Group	arv (single network)	arv (mixture of experts)	p (single network vs. mixture of experts)
Acromegaly, basal	0.51 ± 0.29	0.24 ± 0.13	0.01
Acromegaly, SMS	0.54 ± 0.22	0.21 ± 0.11	0.005
Control, basal	0.14 ± 0.07	0.07 ± 0.03	0.006
Control, fasting	0.12 ± 0.04	0.09 ± 0.06	0.04

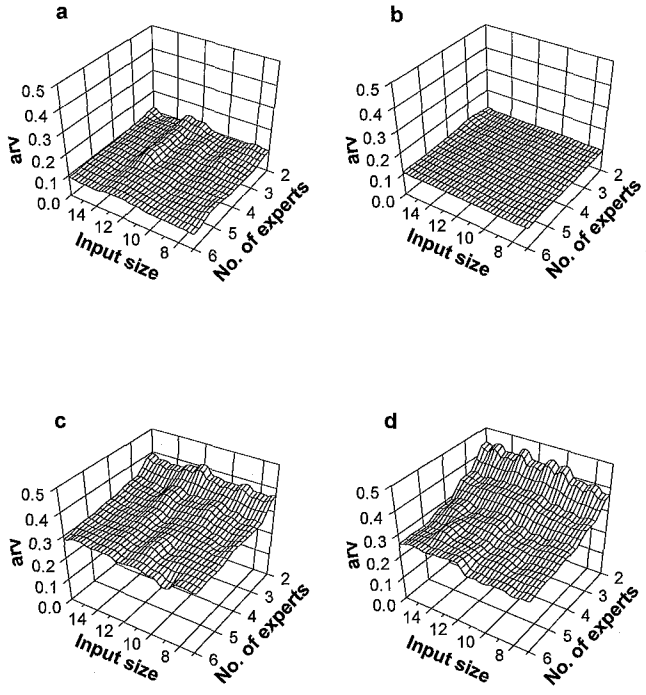

FIGURE 3 One-step-ahead prediction error (arv) as a function of the number of input units and number of experts. Each system was trained and tested to perform a one-step-ahead prediction, using a leave-one-out technique, in which all but one time series were used for training. Finally, the validation time series was used for testing the predictive performance. (*a*) Healthy controls, basal state; (*b*) healthy controls, fasting state; (*c*) acromegalic patients, basal state; (*d*) acromegalic patients, octreotide treatment.

under fasting conditions (Fig. 1 *b*) and acromegalic patients under octreotide treatment (Fig. 1 *d*).

Computing the autocorrelation function and the corresponding decay time for each GH time series, we could significantly separate acromegalic patients from healthy controls in the basal state (Table 1). However, using the autocorrelation function as a discrimination statistic, acromegalic patients under octreotide treatment were not significantly separable from healthy controls under both study conditions, and acromegalic patients under basal conditions were not significantly separable from healthy controls under fasting conditions (Table 1).

TABLE 4 Mixture of experts predictor

Group 1 vs. group 2	arv (mixture of experts)	*p* (group 1 vs. group 2)
Acromegaly, basal vs. Control, basal	0.24 ± 0.13 0.07 ± 0.03	0.01
Acromegaly, basal vs. Control, fasting	0.24 ± 0.13 0.09 ± 0.06	0.02
Acromegaly, SMS vs. Control, basal	0.21 ± 0.11 0.07 ± 0.03	0.01
Acromegaly, SMS vs. Control, fasting	0.21 ± 0.11 0.09 ± 0.06	0.03

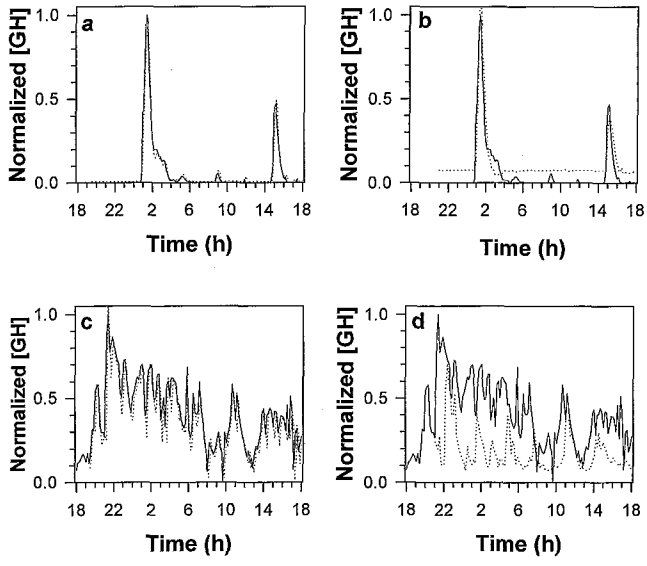

FIGURE 4 One- and ten-step-ahead prediction, using an adaptive mixture of five local experts with nine input units each (nonlinear gating network with three hidden units). ——, Original time series of normalized GH serum concentration; ·····, predicted time series. Healthy control: (*a*) one-step-ahead (10 min) prediction; (*b*) ten-steps-ahead (100 min) prediction; acromegalic patient: (*c*) one-step-ahead (10 min) prediction; (*d*) ten-steps-ahead (100 min) prediction.

Table 2 displays the architecture of the best predictive single feedforward neural network for each group under either study condition. In each subject/patient group, nonlinear networks with hidden units and sigmoidal activation functions demonstrated higher short-term predictability than linear networks without hidden units.

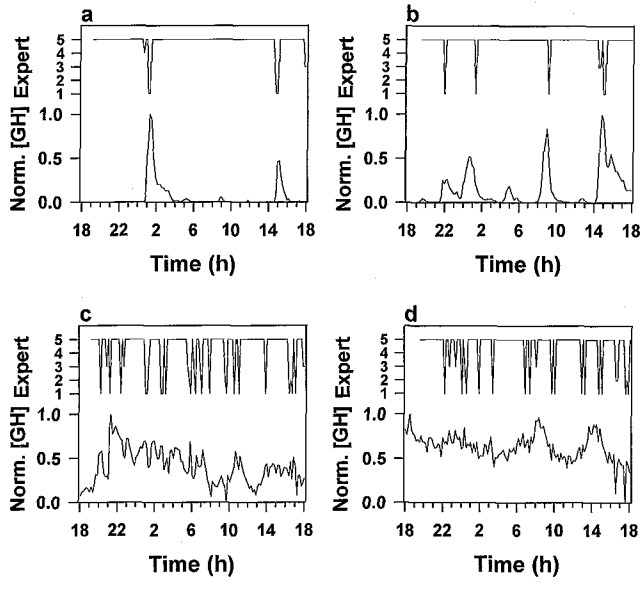

FIGURE 5 Self-organized segmentation of the pulsatile pattern of secretion (SOPUL). Normalized original GH time series and selection pattern of local experts used for prediction. Healthy control: (*a*) basal state; (*b*) fasting state; acromegalic patient: (*c*) basal state; (*d*) octreotide treatment.

TABLE 5 Frequency distribution of local experts

Group 1 vs. group 2	Expert 1 (%)	Expert 5 (%)
Acromegaly, basal	12.7 ± 6.7	82.2 ± 9.7
Control, basal	2.5 ± 1.6	96.0 ± 2.1
	$p = 0.009$	$p = 0.01$
Acromegaly, basal	12.7 ± 6.7	82.2 ± 9.7
Control, fasting	4.4 ± 1.2	94.4 ± 1.8
	$p = 0.02$	$p = 0.02$
Acromegaly, SMS vs.	19.7 ± 9.6	73.0 ± 12.6
Control, basal	2.5 ± 1.6	96.0 ± 2.1
	$p = 0.003$	$p = 0.003$
Acromegaly, SMS vs.	19.7 ± 9.6	73.0 ± 12.6
Control, fasting	4.4 ± 1.2	94.4 ± 1.8
	$p = 0.007$	$p = 0.005$

When we extended the single neural network approach to a modular connectionist system (Fig. 2), healthy controls were easily separable from acromegalic patients by means of different compositions of adaptive mixtures of local experts evaluated to predict best in the respective group (Fig. 3). The best predictive performance of the 24-h GH serum concentration time series in healthy controls was obtained by using an adaptive mixture of five local experts (linear activation functions, no hidden units) with nine input units each (90-min input window) and a gating network consisting of nine input units and three hidden units with sigmoidal activation functions (Fig. 3 *a*), whereas the 24-h GH profile in acromegalic patients was best predicted by using an adaptive mixture of six local experts (linear activation functions, no hidden units) with nine input units each (90-min input window) and a gating network consisting of nine input units and four hidden units with sigmoidal activation functions (Fig. 3 *c*). The one-step-ahead prediction error (arv) was found to be significantly lower in healthy controls under both normal nutrition (Fig. 3 *a*) and fasting (Fig. 3 *b*) than in acromegalic patients with (Fig. 3 *c*) or without (Fig. 3 *d*, summary in Table 4) treatment. This was reflected in correctly predicting the pulsatile features of the 24-h GH profile in healthy controls for 100 min (10 steps, Fig. 4 *b*), in contrast to a significant predictability of only 10 min in acromegaly (Fig. 4 *d*). Compared to a single neural network predictor, the predictive performance of the modular approach was significantly superior for each of the subgroups (Table 3).

The switching between distinct expert networks in the multiple network approach may be used as a self-organized segmentation of the pulsatile secretory pattern (SOPUL) to analyze the temporal GH pattern in healthy controls and acromegalic patients (Fig. 5). The selection pattern of the local experts automatically assigned to a particular case by the gating network easily separates acromegalic patients under both study conditions from healthy controls in the basal state and after fasting (Fig. 5). During the training phase, each expert network specialized to predict certain parts of a GH pulse. Expert 1 focused its performance on predicting the decrease of a hormonal pulse, whereas expert 5 was best in predicting the increase of a GH pulse and the GH baseline. Although experts 2 to 4 contributed significantly to the predictive performance of this modular approach, they were selected in only about 1% of the cases in healthy controls and in about 5% to 8% of the cases in the patient group. The frequency distributions of these most probable experts significantly differ in acromegalic patients from those in healthy controls (Table 5). Expert 1 was chosen in 2.5% to 4.4% in normal controls under basal and fasting conditions, compared to 12.7% to 19.7% in acromegaly under basal conditions and SMS treatment, which separated the controls from acromegaly with high confidence.

DISCUSSION

By computing the autocorrelation function, a well-established tool for time series analysis, the 24-h GH secretory pattern in healthy controls could be separated from that in acromegalic patients only in basal state. Using time series prediction with a single feedforward neural network, we were able to extend the discrimination of the secretory dynamics of GH between healthy subjects and acromegalic patients, despite comparable mean levels under octreotide treatment. We previously used the same analytical approach successfully to predict PTH serum levels and to separate healthy controls from osteoporotic patients for whom such standard methods of time series analysis as computing the mean, the variance, or the power spectrum had failed (Prank et al., 1995). However, such a single neural network predictor has distinct disadvantages when applied to the prediction of time series with nonuniform local dynamics, as described for episodic hormone release in most endocrine systems with phases of secretory bursts and quiescence (Brabant et al., 1992). This may elicit strong interference effects in single feedforward neural networks and result in slow learning and poor generalization. To circumvent these problems we used adaptive mixtures of local experts in the present study. The modular neural network architecture reduces interference by a mixture of several distinct expert neural networks, which are chosen by a gating network to optimally adapt to each training and testing case. By this approach each experimental group was predicted much better than with the single neural network. Each expert specializes in distinct parts of a time series. The local experts selected in the vast majority of cases seem to perform some sort of "edge detection" according to the specialization in the increasing and decreasing phases of a hormonal pulse, without prior knowledge of the form of a pulse or a model of its generation. The switching between local experts may be used to perform a self-organized segmentation of the pulsatile pattern of hormone secretion (SOPUL). We show that in acromegalic patients the expert specialized in predicting the decrease of a GH pulse contributed significantly

more to the overall prediction of the GH time series than in healthy controls. Thus, histograms of selected expert networks may be used to separate temporal rhythms in health and disease.

In contrast to classical approaches for the analysis of pulsatile patterns of secretion (Prank and Brabant, 1994), our SOPUL approach requires no previous knowledge of the physiological characteristics of the system. Therefore, it may better reflect the "true" pattern of endocrine information transfer than the approaches published so far, which all rely on certain assumptions on the nature of a pulse. The approach discussed here may help to expand time series analysis beyond counting frequency and amplitude of hormone pulses.

Recently the question of temporal coding has been raised for the nervous system (Ferster and Spruston, 1995). If a mean rate code is operative, the average number of spikes within a given temporal integration window will carry all of the information about the message. Rearranging spike time patterns without changing the number of spikes should have no effect if neurons encode visual information only by a mean rate code (Barlow, 1972). However, neocortical neurons are capable of precise spike timing (Mainen and Sejnowski, 1995), and temporal coincidences have been observed between pairs of cortical neurons (Abeles, 1991; Engel et al., 1992). Thus temporal coding may be used to increase the efficacy of information transfer. To date, experimental approaches in endocrinology have selectively tested cellular responses to mean rate coding but not temporal coding by pulsatile hormone stimulation (Haisenleder et al., 1992; Knobil, 1980; Schöfl et al., 1993; Shupnik, 1990; Weigle and Goodner, 1986), as all available methods analyze the mean frequency of pulses. Our approach, using multiple adaptive neural networks, may test, from an analytical basis, the potential importance of temporal coding in the endocrine system.

This work was supported by Deutsche Forschungsgemeinschaft under grant Br 915/4-1. SJN and TJS were supported by the Howard Hughes Medical Institute.

REFERENCES

Abeles, M. 1991. Corticonics: Neural Circuits of the Cerebral Cortex. Cambridge University Press, Cambridge.

Barlow, H. B. 1972. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception.* 1:371–394.

Blinowska, K. J., and M. Malinowksi. 1991. Non-linear and linear forecasting of the EEG time series. *Biol. Cybern.* 66:159–165.

Brabant, G., K. Prank, and C. Schöfl. 1992. Pulsatile patterns in hormone secretion. *Trends Endocrinol. Metab.* 3:183–190.

Chang, T., S. J. Schiff, T. Sauer, J. P. Gossard, and R. E. Burke. 1994. Stochastic versus deterministic variability in simple neuronal circuits. I. Monosynaptic spinal cord reflexes. *Biophys. J.* 67:671–683.

Clark, R. G., J. O. Jansson, O. Isaksson, and I. C. Robinson. 1985. Intravenous growth hormone: growth responses to patterned infusions in hypophysectomized rats. *J. Endocrinol.* 104:53–61.

Engel, A., P. König, A. Kreiter, T. Schillen, and W. Singer. 1992. Temporal coding in the visual cortex—new vistas on integration in the nervous system. *Trends Neurosci.* 15:218–222.

Ferster, D., and N. Spruston. 1995. Cracking the neuronal code. *Science.* 270:756–757.

Haisenleder, D. J., G. A. Ortolano, A. C. Dalkin, M. Yasin, and J. C. Marshall. 1992. Differential actions of thyrotropin (TSH)-releasing hormone pulses in the expression of prolactin and TSH subunit messenger ribonucleic acid in rat pituitary cells in vitro. *Endocrinology.* 130: 2917–2923.

Hampshire, J., and A. Waibel. 1989. The meta-pi network: building distributed knowledge representations for robust pattern recognition, technical report CMU-CS-89-166. Carnegie Mellon University, Pittsburgh, PA.

Ho, K. Y., J. D. Veldhuis, M. L. Johnson, R. Furlanetto, W. S. Evans, K. G. M. M. Alberti, and M. O. Thorner. 1988. Fasting enhances growth hormone secretion and amplifies the complex rhythms of growth hormone secretion in man. *J. Clin. Invest.* 81:968–975.

Isgaard, J., L. Carlsson, O. G. Isaksson, and J. O. Jansson. 1988. Pulsatile intravenous growth hormone (GH) infusion to hypophysectomized rats increases insulin-like growth factor I messenger ribonucleic acid in skeletal tissues more effectively than continuous GH infusion. *Endocrinology.* 123:2605–2610.

Jacobs, R. A., M. I. Jordan, S. J. Nowlan, and G. E. Hinton. 1991. Adaptive mixtures of local experts. *Neural Comp.* 3:79–87.

Jordan, M. I., and R. A. Jacobs. 1994. Hierachical mixtures of experts and the EM algorithm. *Neural Comp.* 6:181–214.

Knobil, E. 1980. The neuroendocrine control of the menstrual cycle. *Recent Prog. Horm. Res.* 36:53–88.

Lapedes, A. S., and R. M. Farber. 1987. Nonlinear signal processing using neural networks: prediction and system modeling, technical report LA-UR-87-2662. Los Alamos National Laboratory, Los Alamos, NM.

Lefebvre, J. H., D. A. Goodings, M. V. Kamath, and E. L. Fallen. 1993. Predictability of normal heart rhythms and deterministic chaos. *Chaos.* 3:267–276.

Mainen, Z. F., and T. J. Sejnowski. 1995. Reliability of spike timing in neocortical neurons. *Science.* 268:1503–1506.

Nowlan, S. J. 1990. Competing experts: an experimental investigation of associative mixture models, technical report CRG-TR-90-5. Department of Computer Science, University of Toronto, Toronto, Canada.

Nowlan, S. J. 1991. Soft competitive adaptation: neural network learning algorithms based on fitting statistical mixtures. Ph.D. thesis, CMU-CS-91-126. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. 110–130.

Nowlan, S. J., and G. E. Hinton. 1992. Simplifying neural networks by soft weight-sharing. *Neural Comp.* 4:473–493.

Pawelzik, K., J. Kohlmorgen, and K. R. Müller. 1996. Annealed competition of experts for a segmentation and classification of switching dynamics. *Neural Comp.* 8:340–356.

Prank, K., and G. Brabant. 1994. Estimating thyrotropin secretory activity by a deconvolution procedure. *Methods Neurosci.* 20:377–389.

Prank, K., S. J. Nowlan, H. M. Harms, M. Kloppstech, G. Brabant, R. D. Hesch, and T. J. Sejnowski. 1995. Time series prediction of plasma hormone concentration: Evidence for differences in predictability of parathyroid hormone secretion between osteoporotic patients and normal controls. *J. Clin. Invest.* 95:2910–2919.

Riedel, M., T. Günther, A. von zur Mühlen, and G. Brabant. 1992. The pulsatile GH secretion in acromegaly: hypothalamic or pituitary origin? *Clin. Endocrinol. (Oxf.).* 37:233–239.

Riedel, M., B. Hoeft, W. F. Blum, A. von zur Mühlen, and G. Brabant. 1995. Pulsatile growth hormone (GH) secretion in lean, and obese men: differential metabolic regulation of GH release during energy restriction. *Metabolism.* 44:605–610.

Santoro, N., M. Filicori, and W. F. Crowley. 1986. Hypogonadotropic disorders in men and women: diagnosis and therapy with pulsatile gonadotropin-releasing hormone. *Endocr. Rev.* 7:11–23.

Schiff, S. J ., K. Jerger, T. Chang, T. Sauer, and P. G. Aitken. 1994. Stochastic versus deterministic variability in simple neuronal circuits. II. Hippocampal slice. *Biophys. J.* 67:684–691.

Schöfl, C., G. Brabant, R. D. Hesch, A. von zur Mühlen, P. H. Cobbold, and K. S. R. Cuthbertson. 1993. Temporal pattern of alpha$_1$-receptor stimulation regulate amplitude and frequency of calcium transients. *Am. J. Physiol.* 265:C1030–C1036.

Scott, D. A., and S. J. Schiff. 1995. Predictability of EEG interictal spikes. *Biophys. J.* 69:1748–1757.

Shupnik, M. A. 1990. Effects of gonadotropin-releasing hormone on rat gonadotropin gene transcription in vitro: requirement for pulsatile administration for luteinizing hormone β-gene stimulation. *Mol. Endocrinol.* 4:1444–1450.

Sugihara, G. 1994. Nonlinear forecasting for the classification of natural time series. *Phil. Trans. R. Soc. Lond. A.* 348:477–495.

Sugihara, G., and R. M. May. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature.* 344:734–741.

Tsonis, A. A., and J. B. Elsner. 1992. Nonlinear prediction as a way of distinguishing chaos from random fractal sequences. *Nature.* 358: 217–220.

Waxman, D. J., N. A. Pampori, P. A. Ram, A. K. Agrawal, and B. H. Shapiro. 1991. Interpulse interval in circulating growth hormone patterns regulates sexually dimorphic expression of hepatic cytochrome P450. *Proc. Natl. Acad. Sci. USA.* 88:6868–6872.

Weigend, A. S., and N. A. Gershenfeld. 1993. Time Series Prediction: Forecasting the Future and Understanding the Past. SFI Studies in the Sciences of Complexity, Proceedings Vol. 15. Addison-Wesley, Reading, MA.

Weigend, A. S., B. A. Huberman, and D. E. Rumelhart. 1990. Predicting the future: a connectionist approach. *Int. J. Neur. Sys.* 1:193–209.

Weigle, D. S., and C. J. Goodner. 1986. Evidence that the physiological pulse frequency of glucagon secretion optimizes glucose production by perifused rat hepatocytes. *Endocrinology.* 118:1606–1613.