Estimation of running frequency spectra using a Kalman filter algorithm

D.W. Skagen

Section for Clinical Research, Medical Department A, University of Bergen, Haukeland Sykehus, N-5021, Bergen, Norway

Received May 1987, accepted December 1987

ABSTRACT

A method is suggested for the computation of running frequency spectra from non-stationary oscillations in a long time series. The method is based on an autoregressive model where the coefficients are assumed to vary slowly. The coefficients are updated using the Kalman filter technique. The method is shown to be superior to ordinary autoregressive spectral estimation based on stationary theory in recognizing rapid changes in the frequencies of oscillations.

Keywords: Time series analysis, autoregressive modelling, Kalman filtering

INTRODUCTION

When analysing oscillations in long time series of physiological variables such as heart rate or blood pressure, it may be of interest to see how the dominating frequencies vary with time. The usual strategy is to estimate the spectrum from small segments of the series, and to obtain running spectra by taking consecutive, often partially overlapping time windows. Frequently fast Fourier transforms or autoregressive models have been used to estimate the spectra; however, these methods have in common the basic assumption that the time series is stationary within the time window under consideration. If the purpose is to study time variations in the spectrum, it is desirable to circumvent this assumption. A possible alternative could be to calculate the spectrum from a model with slowly varying time dependent coefficients. We show here how this can be done using the Kalman filter to update the coefficients in an autoregressive model.

THEORY

Autoregressive modelling

Let $x(t):t \in N$ be the tth member of a discrete time series with zero mean value, and let its autoregressive model of order p be written:

$$x(t) = -\sum_{k=1}^{p} \theta_k x(t-k) + \epsilon(t). \quad (1)$$

If $\epsilon(t)$ can be assumed to be a white noise series with uniform variance, it is well known[1] that the power spectrum of the time series $x(t)$ may be expressed as

$$P(\omega) = 2\mathrm{var}(x-\hat{x})/|1 + \sum_{k=1}^{p} \theta_k \exp(-ik\omega)|^2 \quad (2)$$

where $\omega \in [0, \pi]$ is the angular frequency, and $x(t) - \hat{x}(t) = \epsilon(t)$.

This approach to spectral estimation may be a useful alternative to the traditional procedure using the Fourier transform[2,3], in particular because it can be applied to shorter time series, and because $P(\omega)$ is a continuous function of ω.

Several principles for the determination of the coefficients θ_k have been suggested[4]. The simplest is to minimize $\Sigma\epsilon(t)^2$, which leads to the Yule-Walker equations:

$$\begin{pmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ r_1 & r_0 & \cdots & r_{p-2} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ r_{p-1} & r_{p-2} & & r_0 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \cdot \\ \cdot \\ \theta_p \end{pmatrix} = \begin{pmatrix} r_0 \\ r_1 \\ \cdot \\ \cdot \\ r_{p-1} \end{pmatrix} \quad (3)$$

where

$$r_k = \sum_{t=1}^{n-k+1} x(t)x(t-k)/(n-k+1)$$

In order to take variations in the frequency pattern into account, we shall assume that the coefficients θ_k are time dependent. The approach based on maximizing the error variance then becomes impossible, since we are now free to choose coefficients which make $\epsilon(t) = 0$ for every step. Instead, we shall seek coefficients making $\epsilon(t)$ a white noise series which is independent of $x(t)$ and with prescribed variance. We do this by recursive updating of the coefficients for each time step, using the procedure known as the Kalman filter algorithm[5].

Correspondence and reprint requests to Dr D.W. Skagen

0141-5425/88/030275-05 $03.00

Kalman filter theory

Consider a discrete dynamic system with state vector $\Theta(t):t\epsilon N$ of dimension p which develops according to

$$\Theta(t+1) = F\Theta(t) + W(t) \quad (4)$$

Here, F is a $p \times p$ matrix which for our purpose may be considered constant. The vector $W(t)$ is a white noise series with zero mean and with covariance matrix

$$\Sigma = E[WW^T].$$

We shall assume that the state vector Θ itself cannot be observed. However, we may observe an output $x(t)$ which is assumed to be a linear transform of $\Theta(t)$ with added noise $v(t)$. The relation between $\Theta(t)$ and $x(t)$ can be written

$$x(t) = H(t)\Theta(t) + v(t) \quad (5)$$

where $v(t)$ is assumed to be a white noise series with zero mean and variance σ^2, and $H(t)$ must be known.

To apply this model to the problem of estimating the coefficients in the autoregressive model, we let $\Theta(t)$ be the vector of the coefficients $\theta_k:k = 1,\ldots,p$; $x(t)$ be the time series, and $H(t)$ a row vector of the previous p members of $x(t)$, i.e. $H(t) = (x(t-1)\ldots x(t-p))$. Equation 5 is then equivalent to Equation 1. The system matrix F simply becomes the identity matrix. Note that we have no constant term in the autoregression model, which implies that the time series should be high-pass filtered to justify the assumption that $x(t)$ is a zero mean process.

Recursive updating of the coefficients

Suppose that we have an estimate $\hat{\Theta}(t-1)$ for $\Theta(t-1)$, with covariance matrix $C(t-1)$ of the estimation error, i.e. $C = E[(\Theta - \hat{\Theta})(\Theta - \hat{\Theta})^T]$. The best prediction of $\Theta(t)$ based on this information will be

$$\tilde{\Theta}(t) = F\hat{\Theta}(t-1) \quad (6)$$

We then may write

$$\Theta(t) - \tilde{\Theta}(t) = F\Theta(t-1) + W - F\tilde{\Theta}(t-1)$$
$$= F(\Theta(t-1) - \tilde{\Theta}(t-1)) + W \quad (7)$$

Assuming that W is independent of θ, we may readily calculate the covariance matrix A of the prediction error $\Theta - \tilde{\Theta}$ as

$$A(t) = FC(t-1)F^T + \Sigma \quad (8)$$

The best prediction of $x(t)$ from the previous data is

$$\tilde{x}(t) = H(t)\tilde{\Theta}(t) \quad (9)$$

and the variance of $x(t) - \tilde{x}(t)$ can be expressed as

$$s^2(t) = H(t)A(t)H(t)^T + \sigma^2 \quad (10)$$

by a similar argument, if we assume that v and $H(\Theta - \tilde{\Theta})$ are uncorrelated.

The Kalman filter theorem now states that the best linear estimate for Θ, taking all previous observations into account, is given by

$$\hat{\Theta}(t) = F\hat{\Theta}(t-1) + K(t)\,(x(t) - \tilde{x}(t)) \quad (11)$$

where the Kalman gain vector K is given by

$$K(t) = A(t)H(t)^T/s^2(t) \quad (12)$$

Moreover, the covariance matrix C of the estimation error is given by

$$C(t) = A(t) - K(t)K(t)^T s^2(t) \quad (13)$$

A proof of the theorem is outside the scope of this article. It can be found, usually for a more general case, in several texts, for example Davis and Vinter[6].

A heuristic argument can be given by noticing the similarity to a linear regression problem:

$$\Theta - \tilde{\Theta} = K(x - \tilde{x}) \quad (14)$$

The usual estimate for K is

$$K = E[(\Theta - \tilde{\Theta})(x - \tilde{x})^T]\,E[(x - \tilde{x})(x - \tilde{x})^T]^{-1} \quad (15)$$

The last term is the inverse of the variance s^2 of the prediction error for x. Furthermore, put $x = H\Theta + v$ and $\tilde{x} = H\tilde{\Theta}$. Then

$$K = E[(\Theta - \tilde{\Theta})(H\Theta - H\tilde{\Theta})^T]/s^2$$
$$+ E[\Theta - \tilde{\Theta})(Hv)^T]/s^2 \quad (16)$$

Assuming independence between $\Theta - \tilde{\Theta}$ and Hv, the last term in this expression is zero. The first term may be manipulated to give:

$$K = E[(\Theta - \tilde{\Theta})(\Theta - \tilde{\Theta})^T]H^T/s^2 = AH^T/s^2 \quad (17)$$

Furthermore, the covariance matrix of the estimation error $\Theta - \hat{\Theta}$ may be written in terms of the regression coefficient K as:

$$E[(\Theta - \hat{\Theta})(\Theta - \hat{\Theta})^T] = E[(\Theta - \tilde{\Theta})(\Theta - \tilde{\Theta})^T]$$
$$- Ks^2K^T \quad (18)$$

which is the same as Equation 13.

An algorithm for the recursion follows directly from these considerations. Starting with $\hat{\Theta}(t-1)$ and $C(t-1)$, $\hat{\Theta}(t)$ and $C(t)$ are obtained using Equations 8 to 13, with $H(t) = (x(t-1),\ldots, x(t-p))$. The constant σ^2 and the matrix Σ may be specified in advance. This implies that both the stochastic terms W and v in Equations 4 and 5 are given a specified mean (= 0) and variance. For Σ, we give the diagonal entries a uniform value which may be expressed as $\varrho\sigma^2$, and let the other entries be zero. This means that we assume that the

members of the coefficient error vector W in Equation 4 are independent of each other, and that they have uniform variance. The choice of values for ϱ and σ^2 will be discussed later.

To start the recursion, initial estimates for $\hat{\Theta}$ and C must be provided. The simplest way to do this is to calculate the coefficients in an ordinary, stationary autoregressive model from an initial segment of the time series. Then, C and $\text{var}(x - \hat{x})$ can be calculated using ordinary regression theory.

Performance properties

The idea underlying this approach to spectral estimation was to let the coefficients of the autoregressive model vary slowly with time. It may be of interest to evaluate how well the coefficient vectors adapts to non-stationarities, i.e. how the updating process meets the demands for revision of $\hat{\Theta}$. The difference $x - \tilde{x}$ is the error if $\hat{\Theta}$ is assumed to be constant in the next step, while $x - \hat{x}$ is the error remaining after adaptation of $\hat{\Theta}$ to the new value of H and x. We may express the adaptive ability as the relative reduction in error variance due to adaptation:

$$R = (\text{var}(x - \tilde{x}) - \text{var}(x - \hat{x}))/\text{var}(x - \tilde{x}) \qquad (19)$$

The indicator R may be expressed in a simple form. Using Equations 10, 12 and 13, and expressing $\text{var}(x - \hat{x})$ as $\sigma^2 + HCH^T$, we find after some manipulation:

$$R = ((s^2 - \sigma^2)/s^2)^2 \qquad (20)$$

where $\text{var}(x - \tilde{x})$ is expressed as s^2.

The Dependence of R on the parameters σ^2 and ϱ was evaluated numerically. It was found that R and θ were almost exclusively dependent on the ratio ϱ. With increasing ϱ, the values of R increased. If ϱ was kept constant, the choice of σ^2 made virtually no difference when tested over the range 10^2 to 10^{-6}.

A small value for R ($R \sim 0$) imples that $\hat{x}$ is close to $\tilde{x}$, that is $\hat{\Theta}$ remains almost constant. A large $R(R \sim 1)$ implies that $\hat{\Theta}$ is adjusted very efficiently to x, making the error $x - \hat{x}$ small. Hence, the rate of change of Θ, and accordingly the rate of change of the power spectrum, can be tuned by the choice of the parameter ϱ, the ratio between the entries in Σ and σ^2.

Since it was found that $\hat{\Theta}$ was almost independent of the value applied for σ^2, it follows that $\text{var}(x - \hat{x})$ should be largely independent of σ^2. As noted previously, the expressions $\text{var}(x - \tilde{x}) = \sigma^2 + HAH^T$ and $\text{var}(x - \hat{x}) = \sigma^2 + HCH^T$ were deduced under the assumption that $H(\Theta - \tilde{\Theta})$ and $H(\Theta - \hat{\Theta})$, respectively, were uncorrelated to $v = x - H\Theta$. Since these expressions cannot hold for all choices of σ^2, it follows that the assumption that v is uncorrelated both to $H(\Theta - \tilde{\Theta})$ and to $H(\Theta - \hat{\Theta})$ is not valid. However, it also follows that this has only minor impact on the estimates $\hat{\Theta}$.

It will also be apparent that the estimate $HCH^T + \sigma^2$ of the denominator $\text{var}(x - \hat{x})$ in Equation 2 will be proportional to the selected value for σ^2. Since we may write $HCH^T + \sigma^2 = \sigma^2(1 + \sqrt{R})$, the spectra can be normalized by using $1 + \sqrt{R}$ in the denominator instead of the estimated $\text{var}(x - \hat{x})$. In the subsequent calculations this normalization is applied.

In stationary autoregressive spectral estimation, the choice of the number of terms p in the autoregressive model is an important problem[7]. A commonly used criterion is to select the number p such that $\Sigma(x - \hat{x})^2$ is minimized[8]. With the present principle, this problem is less important because the coefficients are allowed to vary. Hence, it should be sufficient to include the number of terms necessary to identify the number of spectral components. If the number of components is q, then $p = 2q + 1$ will be sufficient. Our numerical studies indicate that this also holds true in practice. As a general rule, we have used $p = 10$.

NUMERICAL EXAMPLES

A program was written in Pascal on a HP 217 desktop computer using the algorithm outlined above. Initial estimates for Θ and C were obtained by solving the Yule-Walker equations for the first 60 points in the series, with the order of the model $p = 10$. The $\text{var}(x - \hat{x})$ in this model was used as σ^2, while ϱ was specified in each case to give values of R around 0.5. Spectra were drawn consecutively for each fifth recursion. To make the spectra more readable, they were plotted on a logarithmic scale, and negative logarithms were put to zero.

For comparison, spectra derived using stationary theory were also computed from consecutive time windows of 60 points which were moved 20 points for each estimate. The order p of the model was 10, and the coefficients were computed by solving the Yule-Walker equations directly.

The first example (*Figure 1*) is an ensemble of sinusoids, illustrating the ability of the model to reproduce frequency variations in sinusoids. The equations used for $t < 480$ were:

$$x(t) = 5(\sin(2\pi\omega_1 t) + \sin(2\pi\omega_2 t)) + \zeta(t) \qquad (21a)$$

with $\omega_1 = 0.1$ and $\omega_2 = 0.2$ for $t < 240$ and $\omega_1 = 0.1 + 0.02\,(t - 240)$ and $\omega_2 = 0.2 - 0.01(t - 240)$ for $240 < t < 480$.

For $t > 480$:

$$x(t) = 5\,\sin(2\pi\omega_1(t + 0.1(\sin(2\pi\omega_2 t)) + \zeta(t) \qquad (21b)$$

with $\omega_1 = 0.15$ and $\omega_2 = 0.01$.

The term $\zeta(t)$ was a pseudo-random number sequence with values in $(-1, 1)$.

In the first part with stationary oscillations, the correct frequencies were identified. In the second part, the linear drift in the frequencies was reproduced very precisely. In the third part, the frequency modulation appears as a sinusoidal variation of the frequency. The rapid adaptation to the resetting of the rhythm at $t = 480$ should also be noted. In comparison, the stationary spectra reproduced the linear drift in frequencies fairly well, but failed to reproduce the frequency modulation.

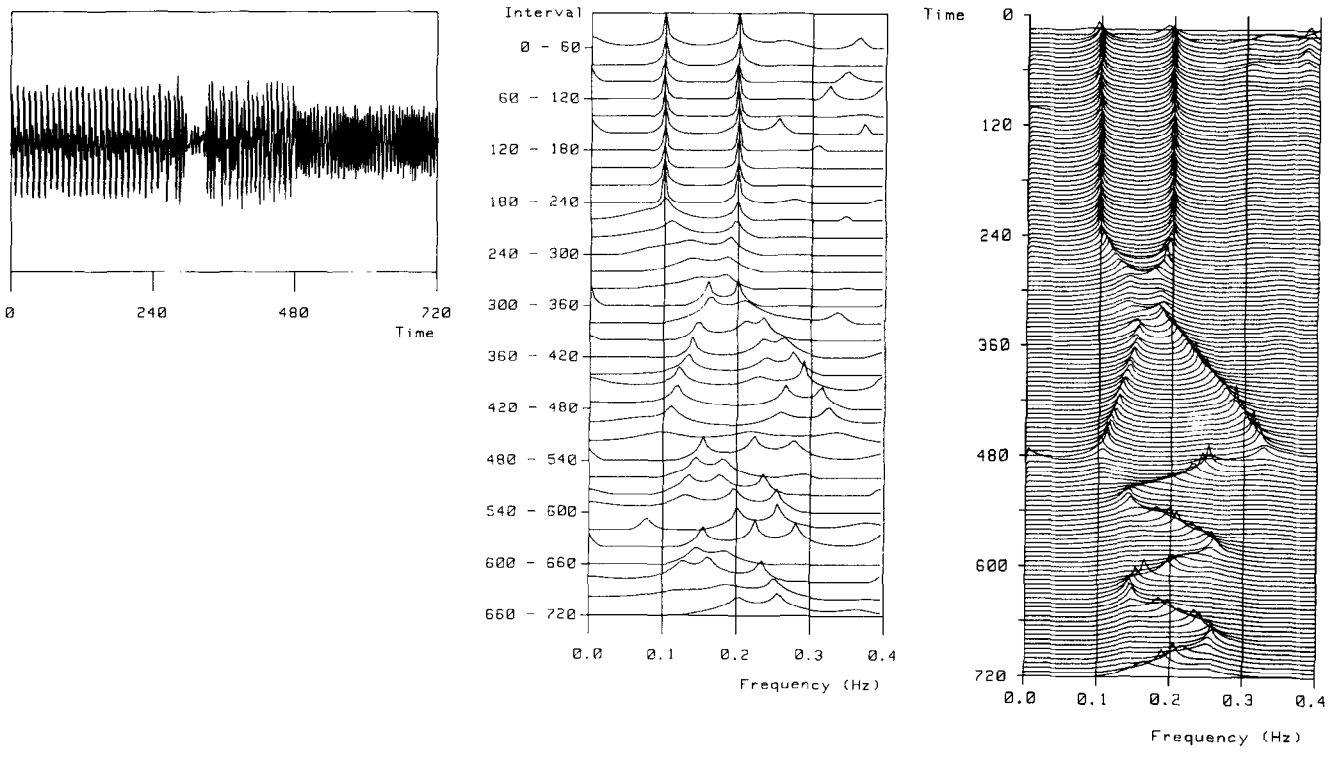

Figure 1 Comparison of power spectra from various sinusoids. **a**, Time series; **b**, log power spectra by the Yule-Walker algorithm; **c**, log power spectra by the Kalman filter algorithm. t = 0 - 240: two stable sinusoids. t = 240-480: two sinusoids with linear frequency drift. t = 480 - 720: a frequency modulated sinusoid

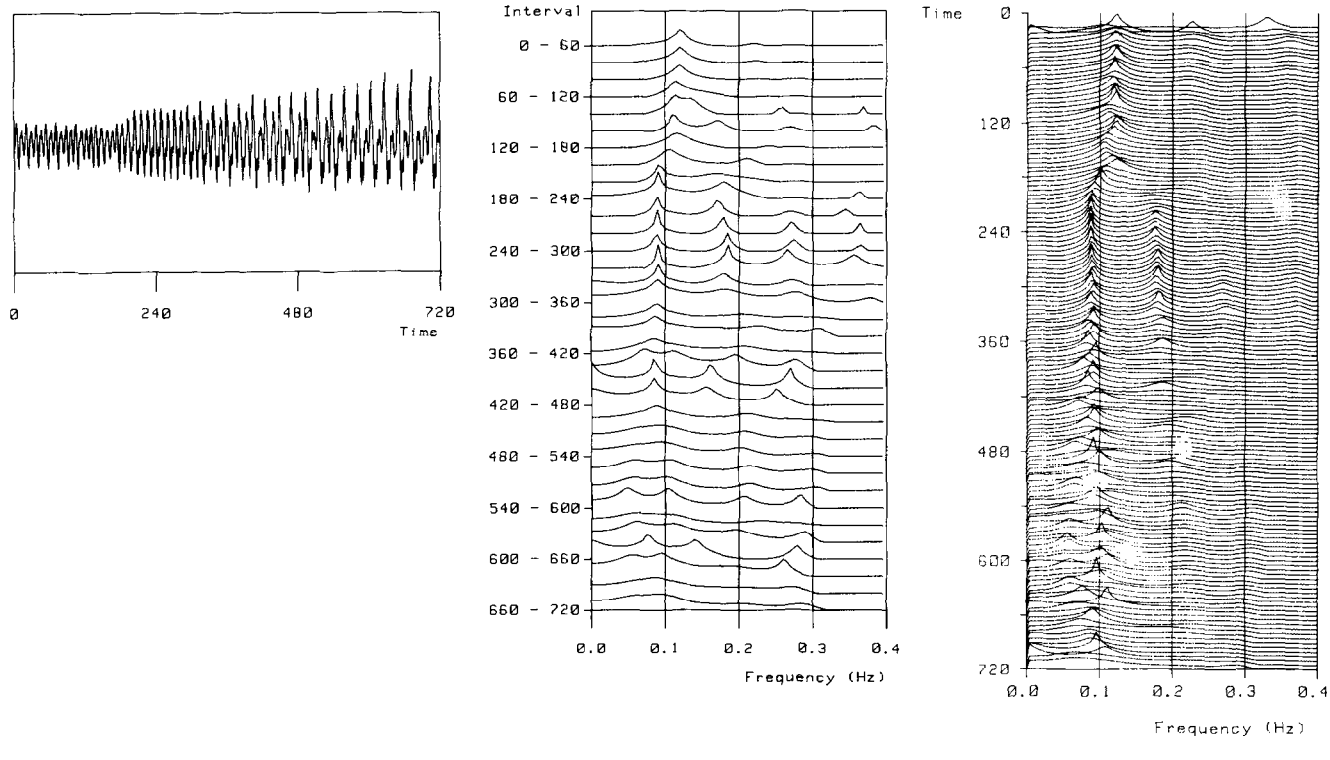

Figure 2 Comparison of power spectra from a nonlinear oscillator with entrainment; **a**, time series; **b**, log power spectra by the Yule-Walker algorithm; **c**, log power spectra by the Kalman filter algorithm

In the second example a nonlinear oscillator was considered. It is described by the differentio-delay equation

$$3dx(t)/dt + 2x(t) = 4x(t - 3) / [[x(t - 3)]^8 + 1] \quad (22)$$

As described by Glass and Mackey[9] this equation has a stable double limit cycle.

For $t > 60$ a forcing term $\alpha(t)\sin(2\pi\omega t)$ was added, with $\omega = 0.09$ and $\alpha(t)$ increasing linearly by $\alpha(t) = 0.001(t - 60)$. When the stability was studied by increasing the amplitude of the forcing term stepwise, we found that the limit cycle degenerated to chaos at approximately $\alpha = 0.02$ ($t = 80$), with gradually changing, slightly irregular oscillations until $\alpha = 0.12$ ($t = 180$) where entrainment took place. The regular rhythm persisted until $\alpha = 0.24$ ($t = 300$). After that, chaos with rapidly changing oscillations, but with a basic frequency near that of the forcing term appeared. This sequence of events was also seen when the amplitude was continuously increased.

As shown in *Figure 2*, when there was no forcing term, the dominating frequency was at approximately 0.12 Hz, which corresponds to the single part of the double limit cycle. The frequency corresponding to the whole limit cycle was not visible, presumably because the two parts of the limit cycle are fairly similar. After entrainment, both the frequency of the forcing term and several higher harmonics appeared. In the region with chaos, the pattern was more variable, with peaks at approximately 0.06 and 0.10 Hz, corresponding to the duration of the main deflations of the time series. The spectra obtained by the Yule-Walker approach gave correct frequency estimates in the regions with stable oscillations, but little information when the rhythm was irregular.

DISCUSSION

We have suggested a method for computing running spectra of long time series by applying an autoregressive model with slowly varying coefficients, which are updated using the Kalman filter algorithm. The principle of Kalman filtering is often suggested as a useful tool in forecasting where it is used to adjust a model for future development as new information becomes available[5]. This approach has also been applied in patient monitoring[10].

Our purpose was slightly different because we were primarily interested in describing the changes that appear in the frequency spectrum. To do this, we used the close relationship between the coefficients in an autoregressive model and the frequency spectrum, and used the updating properties of the Kalman filter to adjust the coefficients in the model as the oscillatory pattern changes. By doing that, we avoided the problems that appear when techniques which assume stationarity are used to observe non-stationarities in the spectra.

We have demonstrated that variable frequencies can be followed very closely by this method. It is also noteworthy that resetting of the oscillatory pattern is recognized very quickly. The recursive procedure itself is sufficiently simple to be applied on-line in many situations. However, it is necessary to assume that the time series has zero mean. We feel that this technique may be a useful tool in the study of physiological variables which are known to oscillate and where variations in the frequency spectrum may be used as a probe to physiological control, such as the heart rate[11], blood pressure[12] and gastrointestinal motor activity[13].

REFERENCES

1. Kay S, Marple S. Spectrum analysis - a modern perspective. *Proc IEEE* 1981; **69**: 1380–419.
2. Kitney RI, Fulton T, McDonald AH, Linkens DA. Transient interactions between blood pressure, respiration and heart rate in man. *J Biomed Eng* 1985; **7**: 217–24.
3. Pagani M, Lombardi F, Guzzetti S *et al.* Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circ Res* 1986; **59**: 178–93.
4. Ulrych TJ, Bishop TN. Maximum entropy spectral analysis and autoregressive decomposition. *Rev Geophys Space Phys* 1975; **13**: 183–200.
5. Harrison PJ, Stevens CF. Bayesian forecasting. *J Roy State Soc B* 1976; **38**: 205–47.
6. Davis MHA, Vinter RB. *Stochastic Modelling and Control.* London: Chapman and Hall, 1985.
7. Linkens DA. Empirical rules for the selection of parameters for autoregressive analysis of biomedical rhythms. *Signal Processing* 1979; **1**: 243–58.
8. Akaike E. Statistical predictor identification. *Ann Inst State Math* 1970; **22**: 203–17.
9. Glass L, Mackey MC. Pathological conditions resulting from instabilities in physiological control systems. *Ann NY Acad Sci* 1979; **316**: 214–35.
10. Allen R. Time series methods in the monitoring of intracranial pressure. Part 1: Problems, suggestions for a monitoring sceme and a review of appropriate techniques. *J Biomed Eng* 1983; **5**: 5–18.
11. Axelrod S, Gordon D, Ubel FA, Barger AC, Cohen RJ. Power spectrum analysis of heart rate fluctuation: A quantitative probe to beat-to-beat cardiovascular control. *Science* 1981; **213**: 220–22.
12. deBoer RW, Karemaker JM, Strackee J. Relationships between short-term blood-pressure fluctuations and heart-rate variability in resting subjects. 1: a spectral analysis approach. *Med Biol Eng and Comput* 1985; **23**: 352–8.
13. Van der Schee EJ, Grashuis JL. Contraction - related, low frequency components in electrogastrographic signals. *Am J Physiol* 1983; **245**: G470–G475.