Adaptive Processing Techniques Based on Hidden Markov Models for Characterizing Very Small Channel Currents Buried in Noise and Deterministic Interferences
Author(s): S. H. Chung, Vikram Krishnamurthy, J. B. Moore
Source: *Philosophical Transactions: Biological Sciences*, Vol. 334, No. 1271 (Dec. 30, 1991), pp. 357-384
Published by: The Royal Society
Stable URL: http://www.jstor.org/stable/55571
Accessed: 19/06/2011 23:31

Adaptive processing techniques based on Hidden Markov Models for characterizing very small channel currents buried in noise and deterministic interferences

S. H. CHUNG[1], VIKRAM KRISHNAMURTHY[2] AND J. B. MOORE[2]

[1] *Protein Dynamics Unit, Department of Chemistry and* [2] *Research School of Physical Sciences & Engineering, Australian National University, Canberra, A.C.T. 2601, Australia*

CONTENTS

SUMMARY

Techniques for characterizing very small single-channel currents buried in background noise are described and tested on simulated data to give confidence when applied to real data. Single channel currents are represented as a discrete-time, finite-state, homogeneous, Markov process, and the noise that obscures the signal is assumed to be white and Gaussian. The various signal model parameters, such as the Markov state levels and transition probabilities, are unknown. In addition to white Gaussian noise, the signal can be corrupted by deterministic interferences of known form but unknown parameters, such as the sinusoidal disturbance stemming from AC interference and a drift of the base line owing to a slow development of liquid-junction potentials. To characterize the signal buried in such stochastic and

Phil. Trans. R. Soc. Lond. B (1991) **334**, 357–384
Printed in Great Britain

deterministic interferences, the problem is first formulated in the framework of a Hidden Markov Model and then the Expectation Maximization algorithm is applied to obtain the maximum likelihood estimates of the model parameters (state levels, transition probabilities), signals, and the parameters of the deterministic disturbances.

Using fictitious channel currents embedded in the idealized noise, we first show that the signal processing technique is capable of characterizing the signal characteristics quite accurately even when the amplitude of currents is as small as 5–10 fA. The statistics of the signal estimated from the processing technique include the amplitude, mean open and closed duration, open-time and closed-time histograms, probability of dwell-time and the transition probability matrix. With a periodic interference composed, for example, of 50 Hz and 100 Hz components, or a linear drift of the baseline added to the segment containing channel currents and white noise, the parameters of the deterministic interference, such as the amplitude and phase of the sinusoidal wave, or the rate of linear drift, as well as all the relevant statistics of the signal, are accurately estimated with the algorithm we propose. Also, if the frequencies of the periodic interference are unknown, they can be accurately estimated. Finally, we provide a technique by which channel currents originating from the sum of two or more independent single channels are decomposed so that each process can be separately characterized. This process is also formulated as a Hidden Markov Model problem and solved by applying the Expectation Maximization algorithm. The scheme relies on the fact that the transition matrix of the summed Markov process can be construed as a tensor product of the transition matrices of individual processes.

1. INTRODUCTION

Measurement of the elementary ionic currents flowing through single channels in the cell membrane has been made possible by the 'giga-seal' patch-clamp technique devised by Hamill *et al.* (1981). A tight seal between the rim of the electrode tip and the cell membrane drastically reduces the leakage current and extraneous background noise, so enabling the resolution of discrete changes in conductances which occur when single channels open or close. With the advent of modern digital processing techniques and the ready availability of computer workstations, it has now become possible to improve the signal-to-noise ratio by more than an order of magnitude and extract information about channel currents which are buried in the noise and which have hitherto been inaccessible. Hints already exist in the literature that some channel currents activated by, for example, glutamate (Jahr & Stevens 1987; Ascher & Nowak 1988; Cull-Candy & Usowicz 1989), intracellular second messengers (Zimmerman & Baylor 1986; Premkumar *et al.* 1990*b*), 5-hydroxy tryptophan (Henderson 1990) and GABA (Premkumar *et al.* 1990*a*) are small relative to the background noise. These microscopic conductance fluctuations occurring in the noise remain largely uncharacterized.

We have devised a technique of identifying and characterizing small channel currents that are obscured by the noise. The methods we propose are based on the assumptions that the onset and offset of transmembrane currents can be represented as a finite-state, first-order, discrete-time, Markov process, and that the noise that corrupts and obscures the signal is stochastic, memoryless (white) and Gaussian. With these assumptions, maximum likelihood estimates of the signal-model parameters and signal statistics contained in the observed set of data are derived. In a previous paper (Chung *et al.* 1990), we provided a detailed account of the theoretical basis for signal processing methods based on Hidden Markov Models (HMM). Its reliability in extracting signals from background noise was then shown, by using both known Markov signal sequences embedded in the noise and noisy channel currents recorded from cultured hippocampal cells. The key to this processing method rests on the Baum–Welch re-estimation formulae, which are in turn based on the re-estimation theorems formulated by Baum *et al.* (1970). Here we have further refined and extended our approach to more effectively learn the discrete state levels, making use of the Expectation Maximization (EM) algorithm (Dempster *et al.* 1977; Titterington *et al.* 1985), of which the Baum–Welch re-estimation formulae are a special case. With the observations recorded in real experimental situations in mind, we have considered the case where the underlying signal is corrupted by, in addition to white Gaussian noise, a deterministic disturbance of known form but unknown parameters. These include periodic disturbances with unknown parameters (frequency components, amplitudes and phases) and a drift of the baseline, the form of which we have assumed can be represented as a polynomial function of time. We illustrate with simulation examples the techniques developed in a companion paper (Krishnamurthy *et al.* 1991*b*) which use the EM algorithm to obtain maximum likelihood estimates of the Markov signal and the parameters of the deterministic interference.

In §2 of the paper, we briefly outline the theoretical basis for the HMM processing methods. From a number of simulation studies, detailed in §3, we show that a Markov signal of amplitude as low as 1/10 of the standard deviation of white Gaussian noise can be characterized accurately with the technique we propose here. In §4, we show methods for dealing with records that are likely to be obtained from real experiments, and which are contaminated, perhaps unavoidably, by deterministic interferences from the electricity mains and a slow drift of the baseline. In §5, we describe a method for decomposing the sum of two or more single-channel currents contributing to total current flow.

2. THEORETICAL BACKGROUND

(*a*) *Overview*

The digital signal processing method we use to characterize the statistics of small channel currents, known commonly as the Hidden Markov Model (HMM) technique, was first formulated by Baum and his colleagues (Baum 1972; Baum & Petrie 1966; Baum *et al.* 1970) and subsequently applied to a variety of numerical estimation problems, including speech processing (Levinson *et al.* 1983), two-dimensional image processing (Besag 1986; Geman & Geman 1984) and biological signal extraction (Chung *et al.* 1990). For further details on statistical inference for Markov processes, the reader is referred to Billingsley (1961).

The HMM signal processing techniques are based on the assumption that the characteristics of the signal we are interested in characterizing and the background noise that obscures the signal are different. It is assumed that the signal can be represented as a discrete-time, finite-state, first-order, homogeneous, Markov process, whereas the noise that obscures the signal is a stochastic process which is white and Gaussian. Although the processing may not be optimal for departures from these assumptions, the methods can quite reliably detect signals that do not conform precisely to a first-order Markov model with added white noise (see Chung *et al.* 1990; Krishnamurthy *et al.* 1991*a*).

Because standard HMM processing techniques are reviewed expansively elsewhere (Juang 1984; Rabiner & Juang 1986; Rabiner 1989; Chung *et al.* 1990; see, in addition, Peters & Walker (1978)), here we give a brief outline of the theory and list the relevant formulae without detailed explanations. We formally define the following notation to be used throughout the paper, indicating matrices and vectors with bold type:

T	Clock time, $1, \ldots, k, \ldots, T$.
N	Number of discrete Markov (conductance) states, $1, 2, \ldots, N$.
Y_T	Observation sequence, $y_1, \ldots, y_k, \ldots, y_T$
$\boldsymbol{Q}$	Transition rate matrix of continuous-time Markov chain
s_k	Discrete-time, first-order Markov chain
$\boldsymbol{q} = \{q_i\}$	Markov states $i = 1, 2, \ldots, N$
$\boldsymbol{A} = \{a_{ij}\}$	Transition probability from state i to j
$\boldsymbol{b} = b_i(y_k)$	Probability of observation y_k given that the signal at time k was at state i, and called 'symbol probability'
$\boldsymbol{\pi} = \pi_i$	Probability that signal is at state q_i at time $k = 1$
λ	Signal model, $\lambda = (\boldsymbol{q}, \boldsymbol{A}, \boldsymbol{b}, \boldsymbol{\pi})$
$\hat{q}_i$, $\hat{a}_{ij}$, $\hat{\lambda}$	Estimates of q_i, a_{ij} and λ
σ_ω	The standard deviation of the white noise

The following abbreviations are used:

HMM	Hidden Markov Model
EM	Expectation Maximization
MAP	Maximum *a posteriori*

(*b*) *Signal model*

The model of channel dynamics proposed by Colquhoun & Hawkes (1977, 1981, 1982) is based on a finite-state, continuous-time Markov process, where the state represents the hypothetical conformational state of the channel macromolecule and the transition rate matrix of the process is denoted by $\boldsymbol{Q}$. The states in this model are aggregated and partitioned into two classes, namely open and closed states. The underlying Markov process is not directly observable but some of its properties can be deduced from the behaviour of single-channel currents. By fitting exponential functions to the observed distributions of open- and closed-time histograms, for example, the number of underlying conformational states, and the rate constants from one conformational state to another, each represented as a continuous-time, two-state, Markov process, can be deduced. We make the following assumptions about the Markov process s_k.

Discrete-time

Time is discrete. The time index k belongs to a set of positive integers. It is convenient to deal with discrete-time Markov processes embedded in noise. Techniques for extracting continuous-time Markov processes from noise are presented in Zeitouni & Dembo (1988), but the mathematics associated with such techniques is relatively difficult involving use of the properties of Wiener processes and Ito stochastic calculus. Because in practice the experimental record we deal with is obtained by sampling continuous-time processes, there is no motivation to add unnecessary mathematical complexity by working with continuous-time processes.

Finite-state

The finite-state assumption implies that for each k, s_k is a random variable taking on a finite number of possible values $q_1, q_2, \ldots, q_N$. Each q_i, where $i = 1, 2, \ldots, N$, is called a state of the process and s_k is termed an N-state Markov chain. We denote the state space $\{q_1, q_2, \ldots, q_N\}$ as $\boldsymbol{q}$. In the context of channel currents, the Markov state s_k represents the true conductance level (or current amplitude) uncontaminated by noise at time k. The observed value y_k contains the signal s_k, random noise w_k and possibly deterministic interferences p_k, such as sinusoidal interferences from electricity mains and baseline drift. Thus, it is assumed that the amplitude of true currents at time k takes on one of N discrete levels, $q_1, q_2, \ldots$ or q_N. The meaning of 'state' in our representation differs from that adopted in the Colquhoun–Hawkes model but is consistent with that used in mathematical literature (Kemeny & Snell 1960; Billingsley 1961). The underlying conformational 'state', which is not directly observable from measurements, does not feature in our scheme.

First-order

The probability of s_{k+1} being in a particular state at time $k+1$, given knowledge of states up to time k, depends solely on the state s_k at time k. That is,

$$P(s_{k+1} \mid s_1, s_2, \ldots, s_k) = P(s_{k+1} \mid s_k).$$

The transition probabilities of passing from state level q_i at time k to state level q_j at time $k+1$, defined as

$$a_{ij} = P(s_{k+1} = q_j \mid s_k = q_i),$$

form a state transition probability matrix $\boldsymbol{A} = \{a_{ij}\}, i = 1, 2, \ldots, N, j = 1, 2, \ldots, N$. Note that $\boldsymbol{A}$ is an $N \times N$ matrix, with its diagonal elements denoting the probabilities of remaining in the same state at time $k+1$, given that the process is found in a particular state at time k. Extension of the processing schemes developed to second- or higher order Markov chains is trivial (see, for example, Krishnamurthy *et al.* (1991 *a*)), although the associated computations are more formidable.

Homogeneous

We assume that the transition probabilities are invariant of time k. It is easy to extend the techniques here to semi-Markov processes, in which the transition probability is a function of the time the process spends in a particular state. A mathematical description of this extension is described in Krishnamurthy *et al.* (1991 *a*). Also to characterize a finite-state Markov chain, we define the initial state probabilities $\boldsymbol{\pi} = \{\pi_i\}$ where $\pi_i = P(s_1 = q_i)$.

We also define the probabilistic function of the Markov chain, known also as the symbol probability, as $\boldsymbol{b} = b_i(y_k)$. In a special case where the noise is Gaussian,

$$b_i(y_k) = \frac{1}{\sqrt{2\pi}\sigma_\omega} \exp\left(\frac{-(y_k - q_i)^2}{2\sigma_\omega^2}\right).$$

Specification of a signal model involves choice of the number of states N, and their amplitudes or state levels. Then, transition probabilities from each of N states to the others, an $N \times N$ matrix, need to be assigned. Moreover, the signal model requires a prior knowledge of the variance of the noise and the initial probability distribution. Throughout, we use the notation $\lambda = (\boldsymbol{q}, \boldsymbol{A}, \boldsymbol{b}, \boldsymbol{\pi})$ to represent the signal model.

Remarks

(i) Theoretically, the open- or closed-time interval histogram tabulated from a first-order, discrete-time, finite-state, homogeneous, Markov chain is distributed according to a single exponential function. The decay-time of this exponential function can be deduced from the transition matrix, or conversely, a Markov chain of any desired statistics can be generated by specifying the transition matrix. For simplicity, we have used such first-order Markov chains to evaluate the processing scheme we have devised. In practice, however, a finite length of data segment does not conform strictly to the first-order Markov statistics. Moreover, interval distributions, especially closed-time distributions, obtained from real channel currents can best be fitted with two or three exponential functions. We see later that the departure from the first-order Markov assumption has little, if any, effect on the performance of the processing scheme (§§ 3*e* and 5*b*). Results of extensive simulations using signal sequences which are not first-order Markovian showed that the processing scheme is insensitive to deviation from this assumption (see, in addition, Chung *et al.* (1990)).

(ii) The transition rate matrix $\boldsymbol{Q}$ of a continuous-time Markov chain and the transition probability matrix $\boldsymbol{A}$ of a discrete-time Markov chain are related by the matrix exponential function

$$\boldsymbol{A} = \exp[\boldsymbol{Q}t],$$

where t is the sampling time. Conversion from one to the other can be easily obtained, for example, from the Kolmogorov–Feller equations (Larson & Schubert 1979).

(*c*) *Baum's re-estimation theorem*

Figure 1 illustrates the processing method schematically. Based on the initial signal model, the observation sequence, Y_T, is processed a pre-determined number of times, and the results of the final iteration provide the characteristics of the signal embedded in the noise. These include the estimates of the time-domain signal sequence s_k, signal amplitudes $\boldsymbol{q}$, transition matrix $\boldsymbol{A}$ and amplitude probability distribution among different states h_T (figure 1 *a*). The processing steps involved in extracting the signal are further elaborated in figure 1 *b*. The forward–backward procedure computes three variables, $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and γ, for each discrete state $q_1, q_2, \ldots, q_N$, for every clock time, $t = 1, 2, \ldots, T$. (The mathematical definitions and physical meanings of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and γ will be stated later.) Thus, for a segment of data of T points, there will be NT values of $\boldsymbol{\alpha}$s, $\boldsymbol{\beta}$s and γs, where N is the number of states (discrete conductance states). From the computed values of these three variables, the conductance states q_i and the transition probabilities a_{ij} from one state to another are re-estimated. By using these re-estimated values, the entire computational procedure is repeated using the same data, as shown schematically in figure 1 *b*.

The rationale behind this iterative procedure rests on the re-estimation theorem formulated by Baum and colleagues (Baum 1972; Baum & Petrie 1966; Baum *et al.* 1970), which states:

$$P(Y_T \mid \hat{\lambda}) \geqslant P(Y_T \mid \lambda). \tag{1}$$

In words, the probability of the observation sequence Y_T, given the re-estimated signal model $\hat{\lambda}$, is greater than or equal to the probability of Y_T, given the previous signal model λ. Thus, the signal sequence estimated using a revised model is more consistent with the data than that estimated using the previous signal model. When the iterative procedure converges, then $P(Y_T \mid \hat{\lambda}) = P(Y_T \mid \lambda)$ and $\hat{\lambda}$ is termed the 'maximum likelihood estimate' of the HMM.

This re-estimation theorem is fully exploited in the signal processing scheme we introduce here. As an illustrative example, we take a fictitious membrane channel that, when activated, shows three current (subconductance) levels, at 50, 120 and 190 fA. Because the small signals are masked by the background noise, the precise subconductance levels are unknown to the observer, as are the transition probabilities from one level to the other. We make

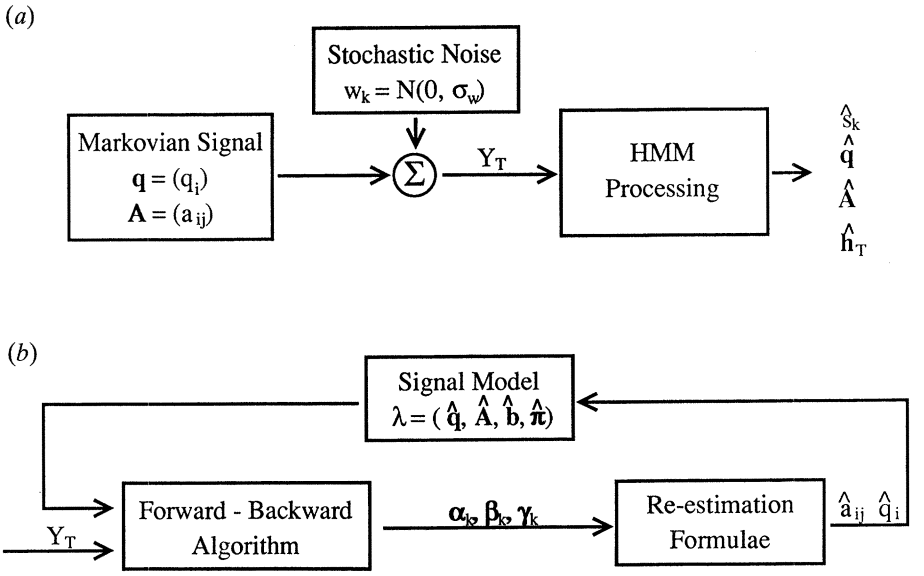

Figure 1. Block diagrams of signal model and the processing method. The model assumed responsible for generation of recorded single channel currents (*a*) and HMM processing method (*b*) are schematically illustrated. (*a*) To a Markov signal sequence with (conductance) states at $q_1, \ldots, q_N$ with the transition matrix $\boldsymbol{A}$, white noise is assumed to be added to give the observation sequence Y_T. The aim of the HMM processing is to obtain the maximum likelihood estimates of the signal sequence s_k, Markov states (conductance states) q_i, transition matrix $\boldsymbol{A}$ and open- and closed-time histograms $\boldsymbol{h}_T$. (*b*) On the basis of the initial signal model λ, the observations sequence Y_T is processed, and three variables α_k, β_k and γ_k are computed for each discrete time k and each Markov state q_i. By using these variables, the parameters of the signal model are revised according to the re-estimation formulae. The entire process is then reiterated.

initial guesses that the four levels q_i (where $i = 1, 2, 3, 4$), including the baseline, are at 0, 100, 200 and 300 fA and assign a 4×4 transition matrix $\boldsymbol{A}$, with all $a_{ii} = 0.88$ and $a_{ij} = 0.04$ (where $i \neq j$). The variance of the noise, calculated from the baseline noise before the channel was activated, is known. The initial parameter estimates are updated after each iteration, the updated parameters are used for the next iteration, and this process continues until convergence. The statistics of the signal sequence extracted from the observed data at convergence are maximum likelihood estimates of the true signal statistics.

(*d*) *Forward–backward procedures*

The HMM techniques are based on two sets of computational steps, known as the forward–backward procedures and Baum–Welch re-estimation formulae. We denote the signal model as λ, which specifies the amplitudes $\boldsymbol{q}$ of N Markov states, the $N \times N$ transition matrix $\boldsymbol{A}$, the standard deviation of the noise σ_ω and the initial distribution probability $\boldsymbol{\pi}$. We define the forward variable $\boldsymbol{\alpha}$ and the backward variable $\boldsymbol{\beta}$ as

$$\alpha_k(i) = P(Y_k, s_k = q_i \,|\, \lambda), \quad \beta_k(i) = P(\bar{Y}_k \,|\, s_k = q_i, \lambda), \quad (2)$$

where Y_k and $\bar{Y}_k$ refer to the past observation sequence $T = 1, 2, \ldots, k$, and the future observation sequence from $k+1$, $k+2, \ldots, T$, respectively. In words, the forward variable $\alpha_k(i)$ is the joint probability of the past and present observation with the present signal in state q_i, given the model λ, and $\beta_k(i)$ is the probability of the future observations given that the present state is q_i and given the model λ.

Recursive formulae for equation (2) are readily calculated using Bayes' Rule as:

$$\alpha_k(j) = \sum_{i=1}^{N} \alpha_{k-1}(i)\, a_{ij}\, b_j(y_k), \quad \alpha_1(j) = \pi_j\, b_j(y_1),$$

$$\beta_k(i) = \sum_{j=1}^{N} a_{ij}\, b_j(y_{k+1})\, \beta_{k+1}(j), \quad \beta_T(i) = 1.$$

The forward variable is calculated in a forward recursion and the backward variable in a backward recursion, thus the variable names and algorithm names.

We denote the probability of being in state q_i at time k, given the observation sequence Y_T and the model λ as $\gamma_k(i)$, i.e. $\gamma_k(i) = P(s_k = q_i \,|\, Y_T, \lambda)$. Then, $\gamma_k(i)$ can be computed from the forward and backward variables using the formula:

$$\gamma_k(i) = \frac{\alpha_k(i)\, \beta_k(i)}{\sum_{i=1}^{N} \alpha_k(i)\, \beta_k(i)}.$$

The maximum *a posteriori* (MAP) state estimate $\hat{s}_k$ at time k is

$$\hat{s}_k^{\mathrm{MAP}} = q_j \text{ where } j = \underset{1 \leqslant i \leqslant N}{\operatorname{argmax}}\, \gamma_k(i) \quad 1 \leqslant k \leqslant T.$$

From the computed $\boldsymbol{\alpha}$s, $\boldsymbol{\beta}$s and γs, the relevant statistics of the Markov signal sequence can readily be calculated. These include the MAP signal sequence estimate, amplitude histograms, transition probability matrix, open-time and closed-time distributions and the likelihood function L_T, as defined by:

$$L_T \overset{\Delta}{=} P(Y_T \,|\, \lambda) = \sum_{i=1}^{N} P(s_T = q_i, Y_T \,|\, \lambda) = \sum_{i=1}^{N} \alpha_T(i). \quad (3)$$

The histogram $\boldsymbol{h}_T$, or probability of dwell-time in each of N discrete states, is calculated as:

$$h_T(i) \triangleq P(q_i \mid Y_T, \lambda) = \frac{1}{T} \sum_{k=1}^{T} \gamma_k(i).$$

(*e*) *Baum–Welch re-estimation formulae*

Because the parameters used for the initial model λ are simply reasonable guesses, the estimates of the signal statistics based on this model will not coincide with the model parameters. For the second and following iterations, we substitute the model parameters $\boldsymbol{q}$ and $\boldsymbol{A} = \{a_{ij}\}$ from the parameters estimated from the previous iteration. The state level q_i and the transition probability a_{ij} are re-estimated from the formulae:

$$\hat{q}_i = \sum_{k=1}^{T} \gamma_k(i)\, y_k \Big/ \sum_{i=1}^{N} \gamma_k(i),$$

$$\hat{a}_{ij} = \sum_{k=1}^{T-1} \xi_k(i,j) \Big/ \sum_{k=1}^{T-1} \sum_{j=1}^{N} \xi_k(i,j),$$

where $\xi_k(i,j) = P(s_k = q_i, s_{k+1} = q_j \mid Y_T, \lambda)$. Note that $\xi(i,j)$ can be calculated recursively as:

$$\xi_k(i,j) = \alpha_k(i)\, a_{ij}(y_{k+1})\, \beta_{k+1}(j)/P(Y_T \mid \lambda).$$

The above two formulae are known as the Baum–Welch re-estimation equations. They can be readily derived by maximizing the likelihood function of the 'fully categorized' data (Titterington *et al.* 1985) given in equation A 1 in the Appendix. For example, solving for $\partial\zeta/\partial\hat{a}_{ij} = 0$ with the constraint that $\sum_{j=1}^{N} a_{ij} = 1$ yields the above re-estimation equation for the transition probabilities. Similarly, solving for $\partial\zeta/\partial\hat{q}_i = 0$ gives the re-estimation equation for Markov state levels.

The proofs that the estimates $\hat{q}_i$ and $\hat{a}_{ij}$ ultimately converge to a local maximum of the likelihood function are given in Baum *et al.* (1970) and Baum & Petrie (1966). Once the estimates $\hat{q}_i$ are obtained, because the noise is Gaussian, the re-estimated symbol probabilities $\hat{\boldsymbol{b}}$ are easily calculated as:

$$\hat{b}_i(y_k) = \frac{1}{\sqrt{2\pi}\sigma_\omega} \exp\left(\frac{-(y_k - \hat{q}_i)^2}{2\sigma_\omega^2}\right).$$

The initial probability distribution $\boldsymbol{\pi}$ can be similarly re-estimated, but in the algorithms implemented here it is not updated. For our application, the statistics of the final estimates are negligibly influenced by the values of the initial condition $\boldsymbol{\pi}$, which are nevertheless needed to initiate the computational procedures.

(*f*) *Elimination of deterministic interferences*

The usual sources of deterministic interferences corrupting the signal are periodic disturbances from the electricity mains and a slow drift of the baseline owing, for example, to the development of liquid-junction potentials. In the Appendix, the re-estimation formulae are listed for estimating the frequency components, phases and amplitudes of the periodic disturbances, and the coefficients of the polynomial functions representing the baseline drift. With these re-estimation formulae, the maximum likelihood estimates of the unknown parameters of the deterministic interferences are obtained, together with the statistics of the Markov signal.

(*g*) *Decomposition of multiple channel currents*

Let u_k and v_k be two homogeneous, first-order, finite-state, Markov processes, with their transition probability matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$. The subscripts 1 and 2 denote the first and second Markov processes which are assumed independent. Consider that the observations y_k are obtained as $y_k = u_k + v_k + w_k$, where w_k is white noise. Here we propose a method for obtaining the transition probabilities and state estimates of u_k and v_k, given the observations y_k. These estimates are obtained by first formulating the problem as a two-vector HMM and then applying HMM processing techniques.

If u_k and v_k have respectively $N_1 = \{p_i\}, i = 1, 2, \ldots, N_1$, and $N_2 = \{q_j\}, j = 1, 2, \ldots, N_2$, possible states, a two-vector Markov process can be defined as $\boldsymbol{S}_k = (u_k, v_k)$, which has $N = N_1 N_2$ states. We denote these states as $Z_1, Z_2, \ldots, Z_N$. Given the observations y_k and the two-vector Markov process $\boldsymbol{S}_k$, we define the transition probabilities $\boldsymbol{A}_T$, the symbol probabilities $b_m(y_k)$ and the initial probabilities π_m as:

$$\boldsymbol{A}_T = a_{mn} = P(\boldsymbol{S}_{k+1} = Z_n \mid \boldsymbol{S}_k = Z_m),$$

$$\boldsymbol{b} = b_m(y_k) = P(y_k \mid \boldsymbol{S}_k = Z_m) = \frac{1}{\sqrt{2\pi}\sigma_\omega} \exp\left(\frac{-(y_k - (p_i + q_j))^2}{2\sigma_\omega^2}\right),$$

$$\boldsymbol{\pi} = \{\pi_m\}; \pi_m = P(\boldsymbol{S}_1 = Z_m),$$

where $Z_m = (p_i, q_j)$.

The forward and backward variables $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ are defined similarly as in equation 2, except that there these variables are now $N_1 \times N_2$ matrices. The $N_1 \times N_2$ matrix $\boldsymbol{\gamma}_k$ is expressed as:

$$\gamma_k(m) = P(\boldsymbol{S}_k = Z_m \mid Y_T, \lambda) = \frac{\alpha_k(m)\, \beta_k(m)}{\sum_{m=1}^{N} \alpha_k(m)\, \beta_k(m)},$$

where $m = (i,j), 1 \leqslant i \leqslant N_1, 1 \leqslant j \leqslant N_2$. MAP estimates of $\boldsymbol{S}_k, u_k$ and v_k are:

$$\hat{\boldsymbol{S}}_k^{\mathrm{MAP}} = Z_m \quad \text{where} \quad m = \underset{1 \leqslant n \leqslant N}{\operatorname{argmax}} \gamma_k(n),$$

$$\hat{u}_k^{\mathrm{MAP}} = p_i \quad \text{where} \quad i = \underset{1 \leqslant g \leqslant N_1}{\operatorname{argmax}} \sum_{j=1}^{N_2} \gamma_k(g,j),$$

$$\hat{v}_k^{\mathrm{MAP}} = q_j \quad \text{where} \quad j = \underset{1 \leqslant h \leqslant N_2}{\operatorname{argmax}} \sum_{i=1}^{N_1} \gamma_k(i,h).$$

The last two formulae stated above are easily justified, because

$$\sum_{j=1}^{N_2} \gamma_k(g,j) = P(u_k = p_g \mid Y_T, \lambda).$$

Thus, $\underset{1 \leqslant h \leqslant N_2}{\operatorname{argmax}} \gamma_k(g,i)$ is the most probable state of u_k

given the observations and the model. The same line of argument applies to the most probable state of v_k. The transition matrix $\boldsymbol{A}_T$ can similarly be estimated using the Baum–Welch re-estimation formulae:

$$\xi_k(m,n) \overset{\Delta}{=} \alpha_k(m)\, a_{mn}\, \beta_k(n)\, b_n(y_{k+1})$$

$$a_{mn} = \frac{\sum_{k=1}^{T} \xi_\kappa(m,n)}{\sum_{k=1}^{T} \gamma_\kappa(m)}.$$

Because u_k and v_k are independent, the transition probability matrix of the vector Markov process $\boldsymbol{S}_k = (u_k, v_k)$ is the Kronecker (tensor) product of the transition matrices of u_k and v_k: $\boldsymbol{A}_T = \boldsymbol{A}_1 \otimes \boldsymbol{A}_2$. Once $\boldsymbol{A}_T$ has been estimated using the above equations, $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ can readily be obtained.

3. SIGNALS BURIED IN IDEAL NOISE

Before the HMM processing scheme can be applied to the analysis of biophysical data, its reliability in extracting known signal statistics and signal estimates buried in background noise needs to be thoroughly tested. For background noise, we used a segment of computer-generated stochastic, Gaussian noise, with zero mean and $\sigma_\omega = 0.1$ pA. Unless stated otherwise, the sampling interval was taken to be 200 μs. The characteristics of this noise closely matched noise from the output of a patch-clamp amplifier (Axopatch 1C) with a 10 GΩ resistor across the input or a patch of membrane sealed to the tip of a glass pipette. We have ascertained that the distribution of this background noise, filtered at 2 kHz (−3 dB, Bessel), is Gaussian, with $\sigma_\omega = 0.099$ pA. The power spectrum of the amplifier noise, determined with the Maximum Entropy Method, was relatively flat up to the Nyquist frequency (see Chung *et al.* 1990). To simulate channel currents, we have generated two-state or three-state, first-order Markov signal sequences of various amplitudes and transition probabilities. Throughout we express the amplitude of the signal in pA or fA and the time axis in ms, but these units can readily be converted to the dimensionless signal-to-noise ratio and sampling points k by taking σ_ω as 100 fA and by assuming a sampling frequency of 5 kHz.

A first-order Markov signal sequence was first added to the noise, and the characteristics of the signal were estimated using the HMM processing scheme. Typically, data segments used for the following simulations contained 20000 points, representing 4 s of real time data, and the computations were done using a workstation computer (Sun IV).

(*a*) *Characterization of a two-state Markov signal*

By using two-state, first-order, Markov signal sequences with transition probabilities $a_{11} = a_{22} = 0.97$, we have ascertained that our processing technique can determine the amplitudes of the hidden signal with an acceptable degree of accuracy.

Figure 2 gives the results of one such simulation. To a segment of background noise (*a*), a binary signal sequence of the amplitude of 25 fA (*b*) was added to give the observation sequence (*c*). We note here again that the length of the data processed was 20000 points, well in excess of the 2000 points illustrated in the figure. The amplitude probability density distributions of the noise and the noisy observations could both be approximated as Gaussian (*d* and *e*), but the mean of the latter, as expected, was shifted by 12.5 fA.

In the estimations, we assumed that the variance of the noise was known and that the underlying signal was two-state Markovian. We processed the noise trace containing no signal and the same noise trace to which a signal sequence was added; the results of simulations are presented respectively in figure 2*f*, *g*. With each successive iteration, the estimated values of q_1 and q_2 converged towards the true values. When the observation sequence contained no signal, the estimated states after 100, 200 and 800 iterations were, respectively, 5.2 fA and −6.8 fA, 1.3 fA and −1.7 fA, and −0.16 fA and −0.31 fA. Because the two estimated states after the last iteration were separated by 1/650 of σ_ω, it can be concluded that the data contained no Markov signal. With the observation sequence containing a signal sequence, the estimated states after 400 iterations were 0.37 fA and −27.6 fA, close to the true signal levels of 0 and −25 fA. The estimates of state levels during the first 20 and last 10 iterations are shown in figure 2*f*, *g* as dotted lines. The logarithm of the likelihood functions $\mathrm{P}(Y_T \mid \lambda)$ are also shown as smooth curves. With successive iterations, the estimates of q_1 and q_2 converged asymptotically to the levels corresponding to the true states. Also, the likelihood function rapidly increased and reached a steady value.

(*b*) *Characterization of a three-state Markov signal*

Although single channels typically assume two conductance states, 'on' and 'off', spontaneous chloride channels (Krouse *et al.* 1986), channels activated by neuroactive amino acids (Jahr & Stevens 1987; Cull-Candy & Usowicz 1989) or intracellular messenger systems (Zimmerman & Baylor 1986; Premkumar *et al.* 1990*a*, *b*), among others, show multiple conductance levels. We show here that our HMM processing scheme is particularly useful in deducing the kinetics of channels which exhibit two or more open state current levels.

We generated a three-state Markov signal sequence, embedded in noise. The current flow through this fictitious single channel, when activated, was quantal in nature, alternating randomly between 'closed' (zero current), 'partially open' (−50 fA) and 'fully open' (−100 fA) states. The task of the HMM processing scheme was to uncover the statistics of the hidden signal. The initial signal model we adopted was one with five discrete current levels (including the baseline), at +100, +105, +110, +115 and +120 fa, with all $a_{ii} = 0.9$ and $a_{ij} = 0.025$. The standard deviation of the noise was correctly specified to be 0.1 pA. With this initial model, the sequence of data was processed 50 times.

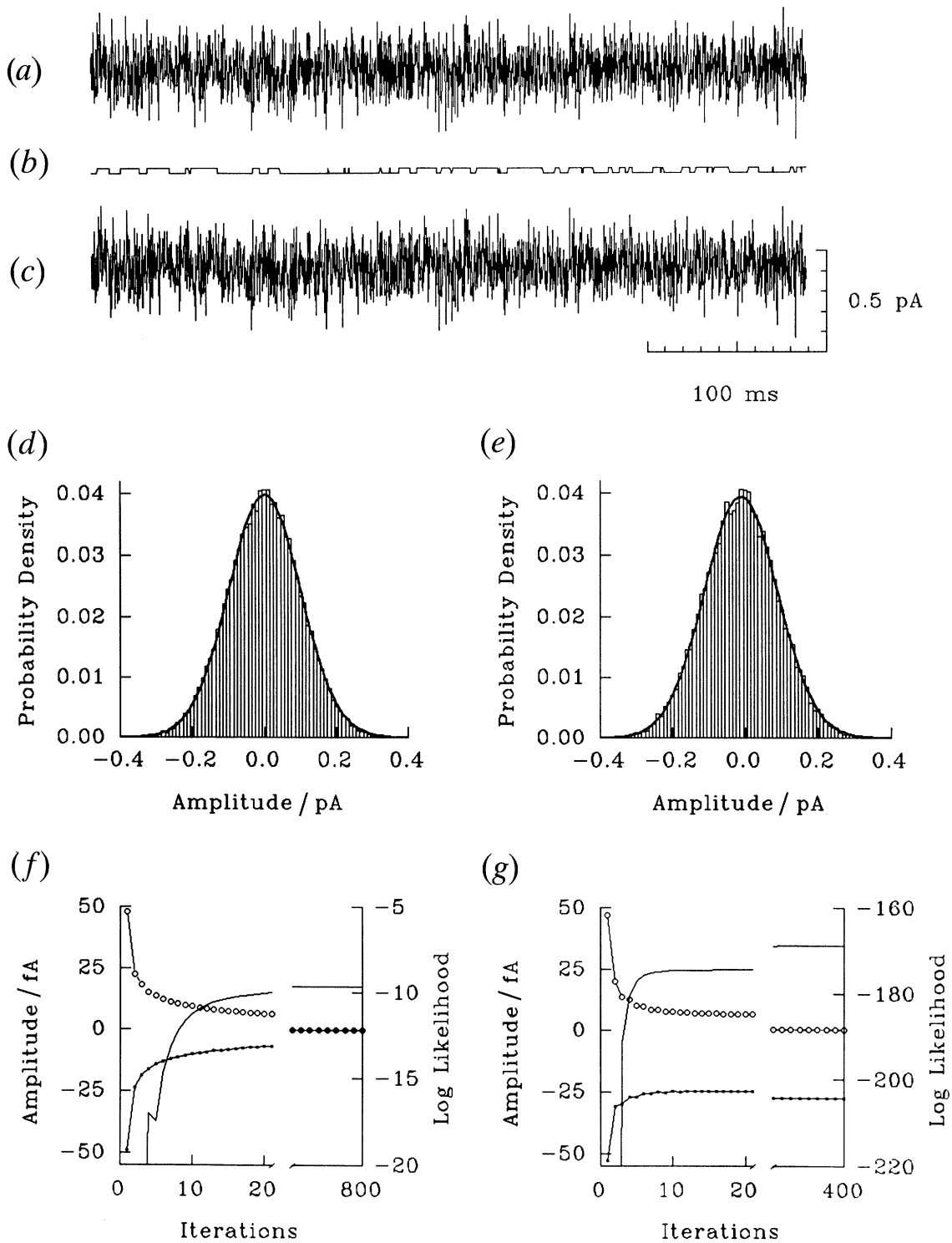

Figure 2. Characterization of a binary Markov process. In this and subsequent figures, unless stated otherwise, the number of data points processed was 20000. Traces of the first 2000 points are plotted for: (*a*) the noise with $\sigma_\omega = 0.1$ pA, (*b*) a two-state Markov signal of amplitude 25 fA, generated according to the transition probabilities $a_{ii} = 0.97$, and (*c*) the signal embedded in the noise. The distributions of the noise (*d*) and the signal buried in the noise (*e*) could be fitted with Gaussian curves (mean $\pm\ \sigma_\omega$: -0.23 ± 100.2 fA for (*d*) and -12.45 ± 101.1 fA for (*e*). The noise alone and data containing the signal sequence were processed, and the estimates of the signal amplitude are plotted in (*f*) and (*g*) against successive iterations. It is assumed that the signal is known to be a two-state Markovian. The initial guesses used for the signal amplitudes and transition probabilities were, respectively, +100 and −100 fA, and $a_{ii} = 0.9$. When the data contained no signal, the estimated amplitudes of the two states coalesced, giving the identical values (*f*). When the data containing the signal was processed, the estimated amplitudes of the two Markov states after 400 iterations were within 10% of the true values (*g*). Open and closed circles represent the amplitude estimates of the open state and of closed state. The solid lines drawn in (*f*) and (*g*) are the logarithm of the likelihood function.

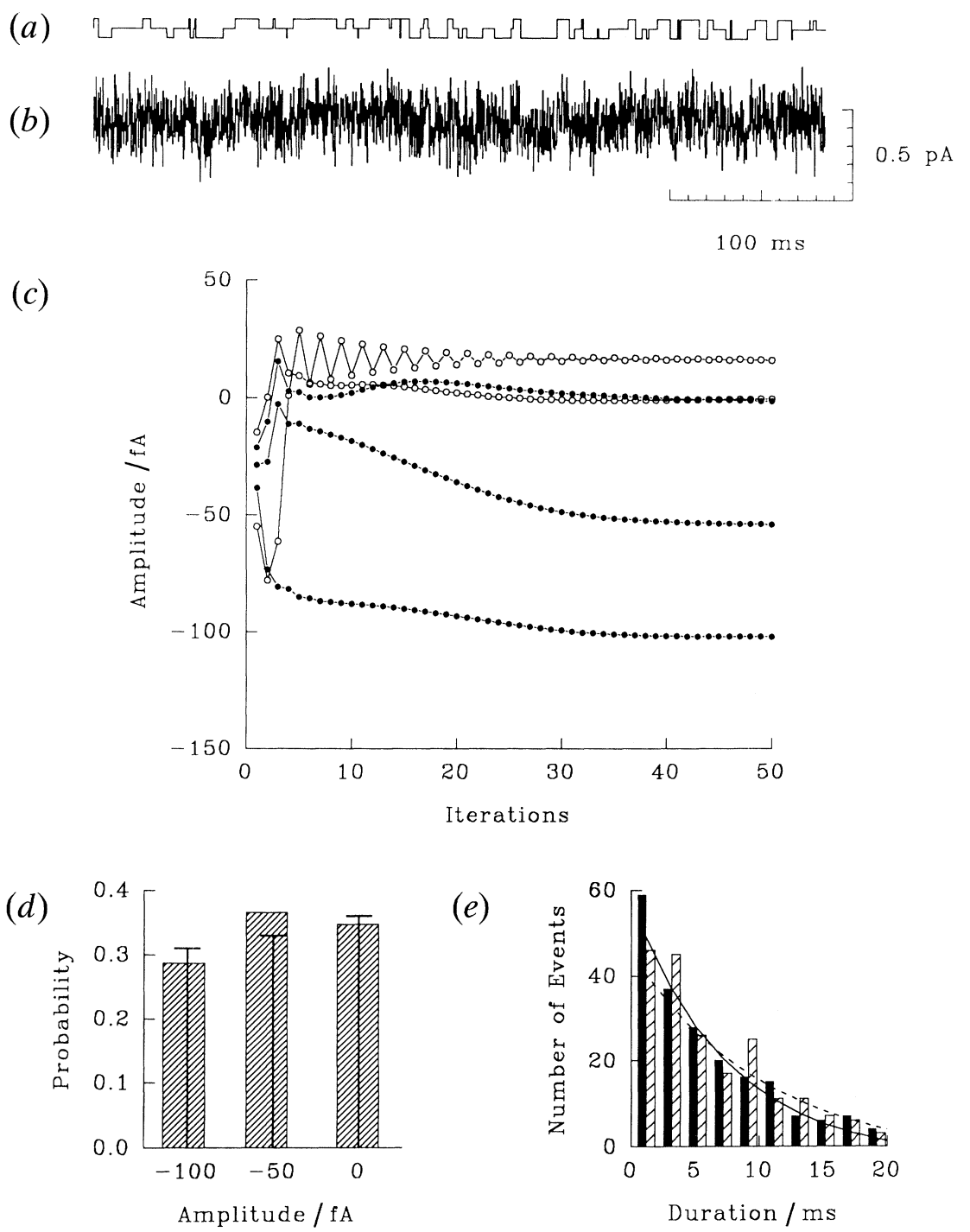

Figure 3. Characterization of a three-state Markov process. A three-state Markov signal, a portion of which is shown in (*a*), was generated according to the transition matrix, $a_{ii} = 0.97$, $a_{ij} = 0.015$. The amplitudes of the three states were 0, −50 and −100 fA. The Markov signal embedded in noise (*b*) was processed, erroneously assuming that there were five states. The estimated amplitude of each of the five states is plotted against successive iterations in (*c*). The estimated amplitude of one state was incorrectly identified as +16.5 fA (open circle), but the probability of the signal being in this state was given as near zero. The estimated amplitudes and probabilities of being in each of the three states are shown in (*d*) in the form of a bar graph. The amplitudes and the relative proportions of dwell time of the original signal are indicated as solid lines. Shown in (*e*) are the open-time distributions of the partially open state (solid line) and the fully open state (broken line), deduced from the estimated transition matrix. The bars show the actual open-time distributions of the partially open state (filled) and fully open state (hatched) of the original signal.

Sample segments of the signal and of the noise containing the signal are shown in figure 3*a*, *b*. The estimated amplitudes of the five states attained steady values by the 30th iteration and remained unchanged during the remaining 20 iterations (figure 3*c*). Two of the five states coalesced to the baseline level, giving estimated amplitudes of −0.95 and −0.06 fA. The estimated signal amplitudes of the two other states at the final iteration were −54 and −102 fA. The estimated amplitude of the remaining state first

oscillated and then reached a steady value of + 16.5 fA. That this last state was a false state could easily be ascertained from a calculation of the probability of the signal being in this state, namely 2×10^{-6}. The probabilities that the original signal was at 0 fA (the baseline), −50 fA and −100 fA are depicted in figure 3*d*, being 0.332, 0.355 and 0.313, whereas the corresponding estimates from the HMM processing scheme were 0.346 (4% error), 0.366 (3% error) and 0.287 (9% error), respectively.

The HMM processing scheme also provided the estimated transition matrix which, omitting the false state and aggregating the two baseline states, reads:

$$\boldsymbol{A} = \begin{pmatrix} 0.975 & 0.009 & 0.016 \\ 0.190 & 0.968 & 0.013 \\ 0.013 & 0.026 & 0.961 \end{pmatrix}$$

where the first, second and third rows refer to 0 fA (the baseline), the −50 fA state and the −100 fA state, respectively.

From this matrix, the distributions of open- and closed-time histograms, the mean open and closed durations and the relative likelihood of transition from one state to the other can be calculated. Theoretically, open-time (or closed-time) distributions of a signal sequence generated by such a matrix should be of the form:

$$P_i(d) = a_{ii}^{d-1}(1-a_{ii}), \tag{4}$$

whereas the mean open (or closed) duration are:

$$E[d] = \sum_{d=1}^{\infty} dP_i(d) = (1-a_{ii})^{-1}.$$

In the above equations, d is expressed in terms of digitized points. By using the estimated transition matrix and equation (4), we computed the open-time distributions and plotted these in figure 3*e*. The solid line is the estimated open-time distribution at −50 fA level, whereas the broken line is that at −100 fA level. Superimposed on the figure and shown in the form of bars are the actual open-time distributions tabulated from the original signal. The filled and hatched bars are the open-time distribution at −50 fA level and −100 fA level, respectively. Because of the short data segment used for the analysis, the open-duration distribution of the original signal deviates considerably from a smooth exponential function. Nevertheless, the curves calculated from the maximum likelihood estimate of the transition probability matrix closely approximate the original signal characteristics.

In summary, the processing scheme estimated with an acceptable degree of accuracy the number of states, their conductance levels, the relative proportion of times spent in each state, mean open- and closed-time durations, open- and closed-time histograms and transition probabilities from one state to the other. If finer accuracy is required, larger sample path lengths would be necessary, but these could then be subject to departures from stationarity of the signal and baseline drift. Also, we could have increased the number of iterations at the expense of computational cost, but with diminishing returns.

(*c*) *Identification of small signals*

What is the smallest signal our processing methods can reliably characterize? The answer to this question depends largely on what we regard as the acceptable degree of error as well as the computational cost we are willing to bear. The smaller the hidden signal is, the larger the uncertainty becomes and the more iterations are needed to achieve a desired accuracy. In general, we find that the estimation errors tend to be acceptably small when the amplitude of the signal is larger than 1/20 of σ_ω, or 5 fA. We illustrate here how the performance of the HMM processing scheme gradually deteriorates as the amplitude of the hidden signal decreases.

A two-state Markov signal sequence embedded in Gaussian noise was analysed. The amplitudes of the binary signals were −20, −15, −10 and −5 fA; short segments are displayed in figure 4*a–d*. The true probabilities of the dwell-time of the signal in the two states were 0.492 (baseline) and 0.508, whereas the noise (figure 4*e*) to which the signal was added, as before, had σ_ω of 100 fA. The estimates obtained are summarized in the form of histograms in figure 4*f–i*. The largest discrepancy between the estimated and true dwell time was about 10% (figure 4*g*). The estimated levels of signals, when the true signal levels were at 0 and −20 fA, were 0.6 and −21.3 fA (figure 4*f*). As the signal amplitudes were reduced in steps of 5 fA, we obtained estimates of: 1.4 and −15.2 fA (figure 4*g*), 0.1 and −9.8 fA (figure 4*h*) and −1.1 and −4.3 fA (figure 4*i*). From these and a number of other simulation results, we conclude that the estimation errors, when a binary signal is separated by 5 fA (1/20 of σ_ω) or less, are unacceptably large and thus the characterization of the signal statistic is beyond the resolution limit of our HMM processing technique.

We have made similar analyses using three-state Markov signal sequences, and the results of these simulations are summarized in figure 5. The signal amplitudes were reduced successively from −80 and −40 fA to −20 and −10 fA (figure 5*a–d*). The probabilities of dwell-time in the baseline, partially open and fully open states were, respectively, 0.33, 0.36 and 0.31, and the noise used had the same statistics as before (figure 5*e*). In figure 5*f–i*, the results obtained after 800 iterations are summarized in the form of histograms. For the signal sequence, whose states are separated by 40 fA, the estimated state levels and their probabilities of dwell-time closely approximate the true values (figure 5*f*). As the signal amplitude decreased, the magnitudes of errors increased. The estimates of the dwell-time probabilities in the intermediate state were the least reliable, whereas the estimations of signal amplitude were somewhat more reliable (figure 5*g–i*).

(*d*) *Effects of the mean signal duration*

The reliability of the HMM processing scheme increases as the mean duration of the signal becomes longer. This is in part due to the fact that errors in the signal sequence estimation tend to occur at the

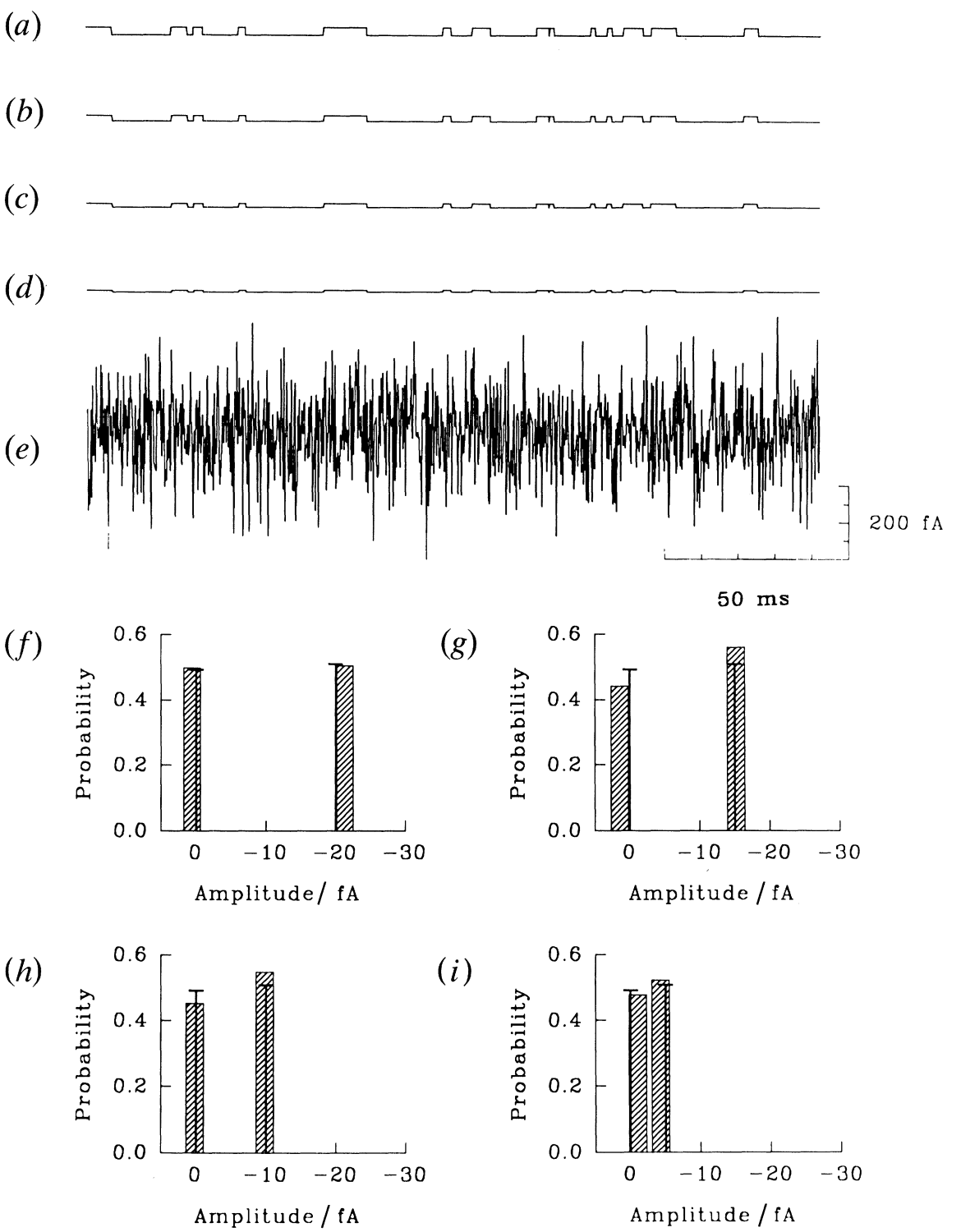

Figure 4. Identification of small binary Markov signals. The amplitude of a binary Markov signal, generated as in figure 2, was progressively reduced from −20 fA to −5 fA in steps of 5 fA (*a*, *b*, *c*, *d*). A 1000 point segment of signal with additive noise is shown in (*e*). The results of the estimated amplitudes and relative proportions of dwell-time in each of the two states, after 400 iterations, are presented as bar graphs (*f*–*i*). The solid lines accompanying the hatched bars are the correct amplitudes and dwell-time probabilities. The initial guesses for the amplitudes of the two state signal were +100 and −100 fA, with the transition probability $a_{ii} = 0.9$. The estimated separations between the closed and open states were −21.9 (true value, −20) fA, −16.6 (−15) fA, −9.9 (−10) fA and −3.2 (−5) fA.

transitions, giving false alarms or misses (Chung *et al.* 1990). Here we show, however, that the overall characterization of the signal statistics is relatively unaffected by the signal durations.

Three-state Markov signal sequences of various mean durations were generated (figure 6*a*–*d*) and added to Gaussian noise of the same variance as used in the previous sections. The three states were separated by 50 fA (0, −50 and −100 fA). The mean duration of the signal was varied from about 6.7 ms ($a_{ii} = 0.97$; figure 6*a*) to 1.3 ms ($a_{ii} = 0.85$; figure 6*d*). The estimates of the probabilities of dwell-time and the

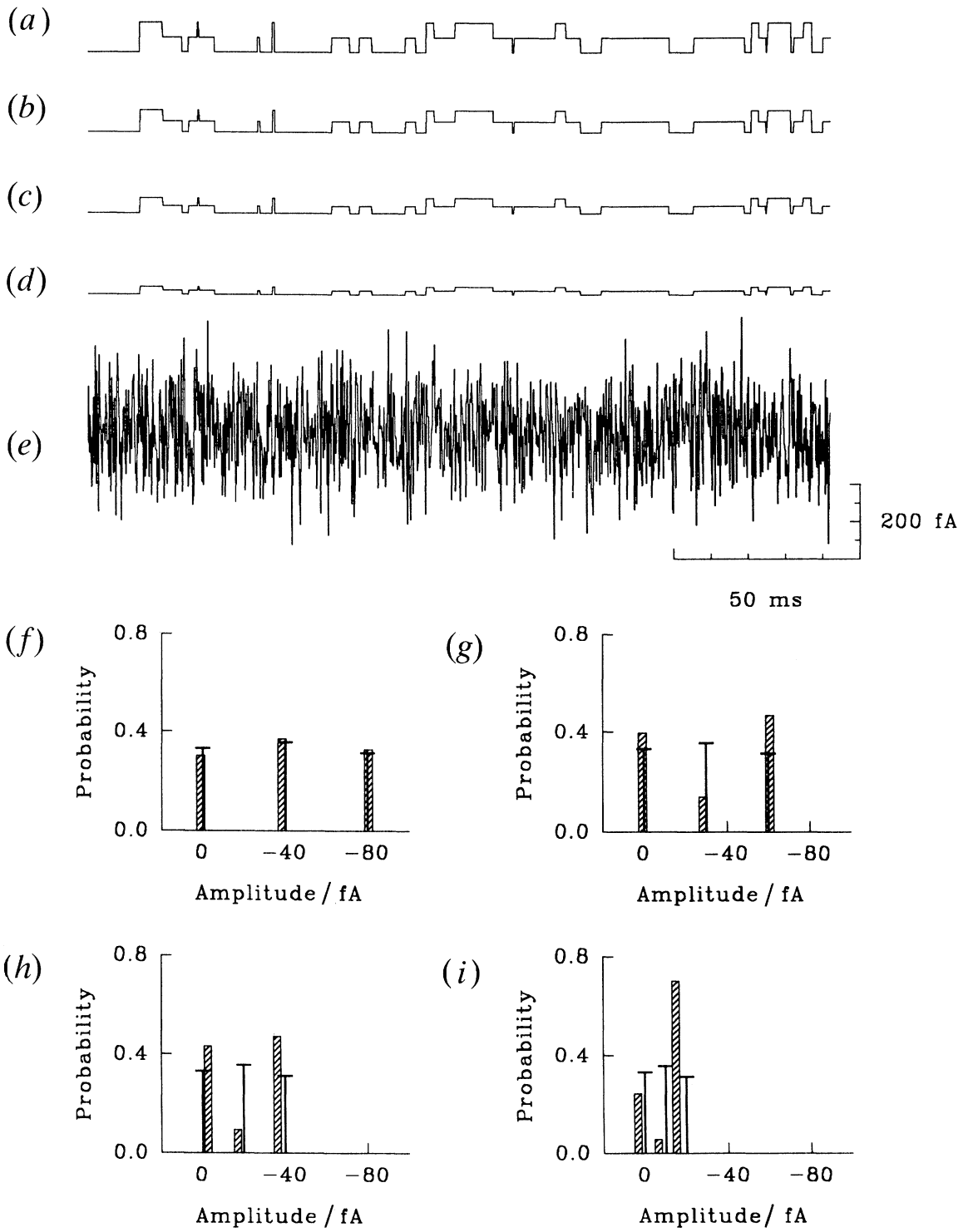

Figure 5. Identification of small three-state Markov signals. Markov signal sequences, three-states equally spaced, were generated as in figure 3. The amplitude separation between the states was 40 fA (*a*), 30 fA (*b*), 20 fA (*c*) and 10 fA (*d*). Each signal sequence was first added to Gaussian noise (*e*) and then processed. For the initial model parameters, the three levels were assumed to be at +25, +50 and +75 fA, and the transition probability $a_{ii} = 0.9$. The results shown in the forms of bar graphs (*f–i*) were obtained after 800 iterations. The estimated amplitudes in fA were: for (*f*) +1.1 (0), −38.5 (−40), −80.6 (−80); for (*g*) +0.9 (0), −28.4 (−30), −60.8 (−60); for (*h*) −2.6 (0), −17.0 (−20), −36.3 (−40); and for (*i*) +3.45 (0), −6.3 (−10), −15.1 (−20). The estimation errors tended to be larger for a three-state Markov process than those for a two-state signal.

signal amplitudes became less accurate as the mean signal duration decreased (figure 6*e–h*). The largest error in the amplitude identification was 7 fA (figure 6*h*), while the probability of dwell time in the fully open state was estimated to be 0.27 as opposed to the true value of 0.32 (figure 6*g*). From the estimated transition probability matrix, we computed the mean duration at each of the three states and compared these with the correct mean duration tabulated from the original signal sequence, as shown in figure 6*i*, where the estimates are plotted against the true values. In general, the mean durations tended to be over-

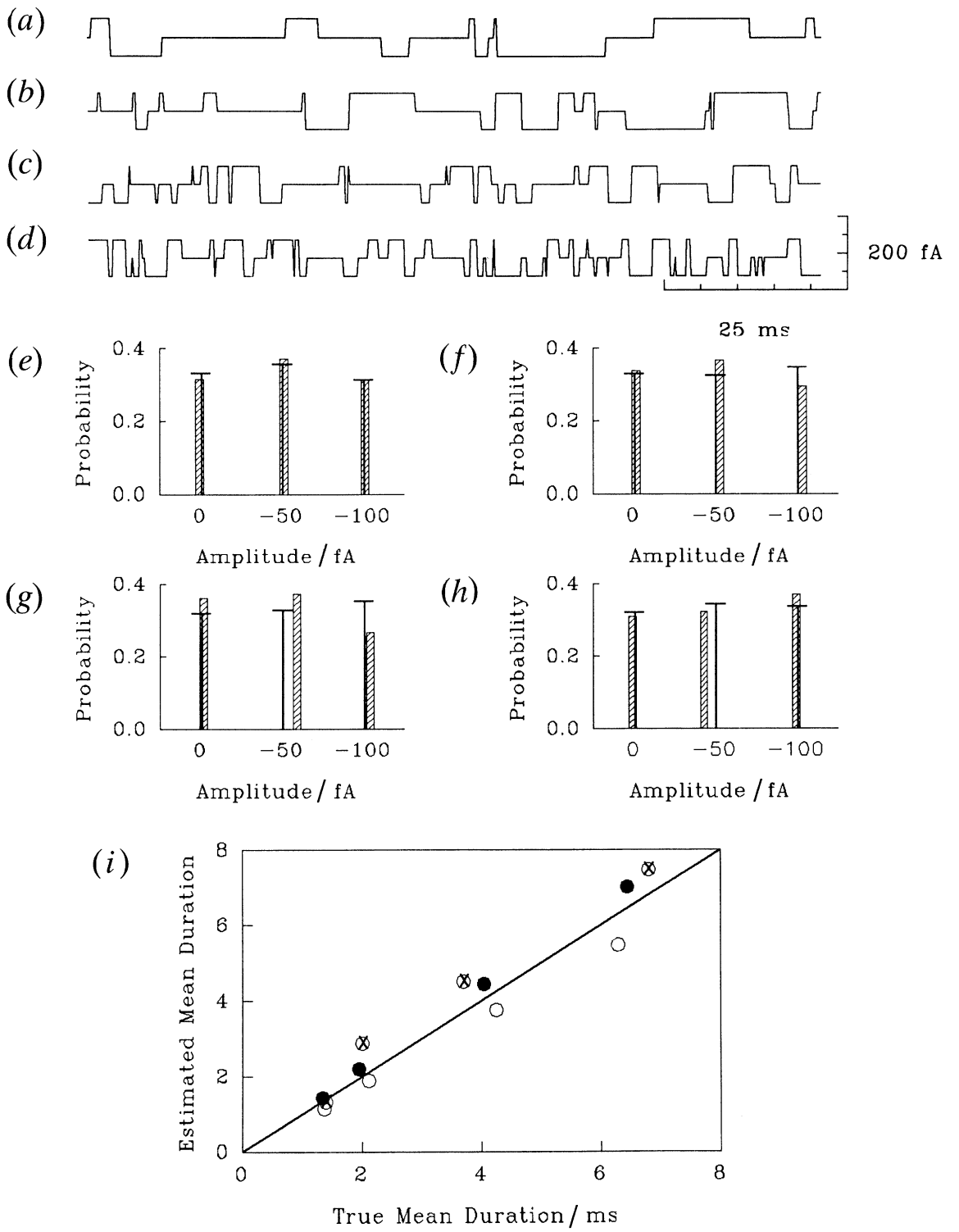

Figure 6. Effects of the mean duration of the signal on estimation errors. First-order, three-state Markov signal sequences of a fixed amplitude but varying mean durations were generated and then added to noise. The transition probabilities a_{ii} used for generating the segments shown in (*a–d*) were, respectively, 0.97, 0.95, 0.90 and 0.85. The expected mean durations of all three states for these segments are 6.7 ms, 4.0 ms, 2.0 ms and 1.3 ms. The states were 0, −50 and −100 fA. The data were iterated 400 times, using the initial guess of the signal amplitudes as +25 fA, +40 fA and +55 fA. The estimated amplitudes and dwell-time probabilities, corresponding to the signals shown in (*a*), (*b*), (*c*) and (*d*) are presented in (*e*), (*f*), (*g*) and (*i*), respectively. The estimated amplitudes in fA were: for (*e*), +0.8, −50.9, −100.9; for (*f*), −1.3, −52.7, −103.5; for (*g*), −1.5, −58.7, −104.1; for (*h*), +1.0, −42.7, −99.4. The correct mean durations of the original signals, calculated from the 20000 point signal sequences, are plotted against the estimated mean durations in (*i*). The filled circles, crossed circles and open circles represent the closed, partially open and fully open states, respectively.

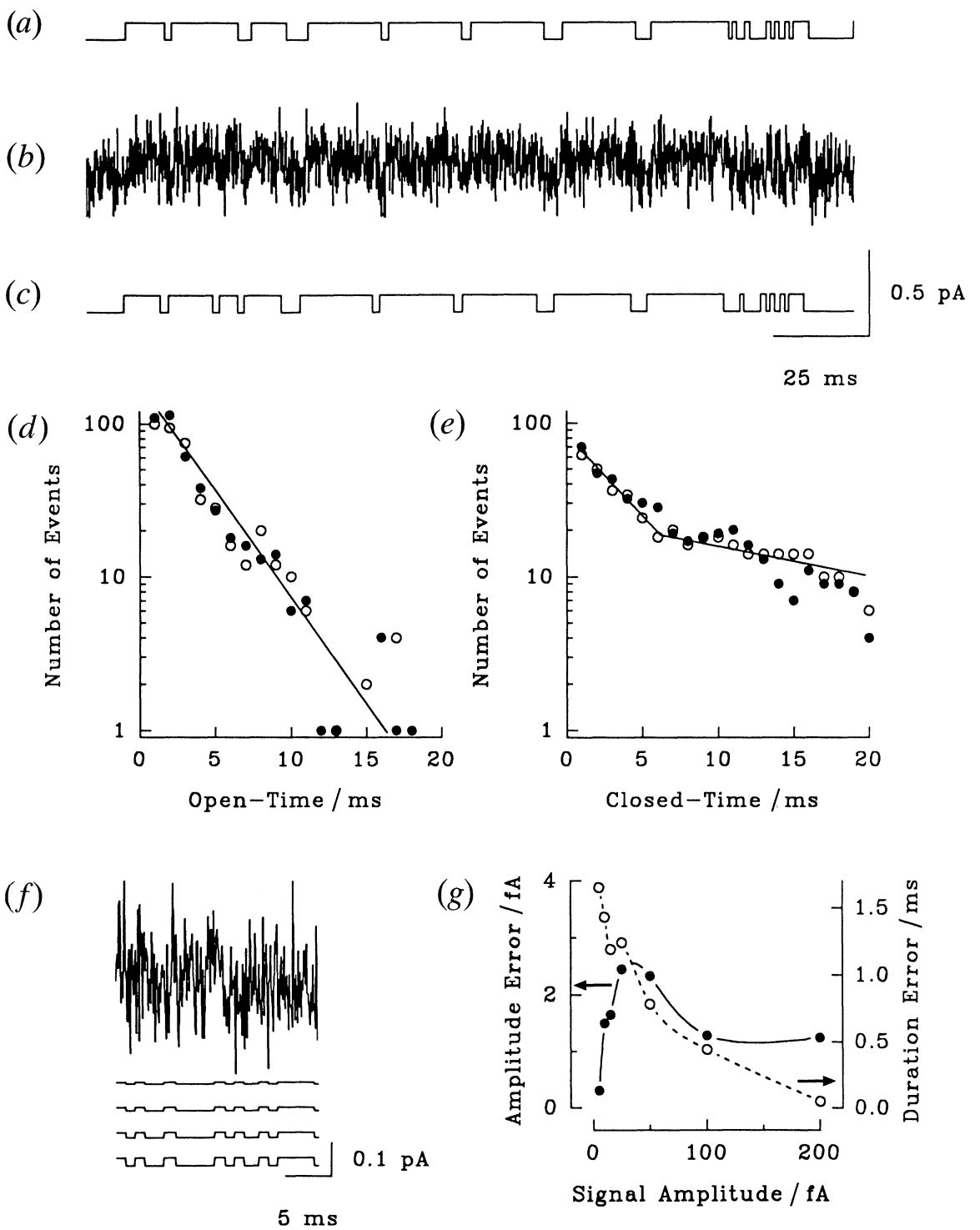

Figure 7. Characterization of a signal sequence that is not first-order Markovian. The signal sequence of two exponential closed-time interval distribution was generated and processed after embedding it to the noise. The amplitude of the segment of the original signal (*a*) was 0.1 pA. The signal was then added to the noise with σ_w of 0.1 pA (*b*), from which the MAP estimate of the signal (*c*) was obtained. In this and all the subsequent figures, the sampling frequency was assumed to be 10 kHz. From a 47000 point segment of the estimated signal sequence, the open- and closed-time interval histograms were constructed (filled circles in *d* and *e*). Superimposed on the graphs, plotted in semi-logarithmic scales, are the interval distributions of the original signal (open circles). Displayed in (*f*) are short segments of the noise, and signals with amplitudes of 5 fA, 10 fA, 15 fA and 25 fA. By using 20000 points of these signals contained in the noise, the amplitudes and mean open-times were estimated. The magnitudes of estimation errors (estimated values − true values) are plotted in (*g*) against signal amplitude. The filled circles refer to the estimation errors of the amplitude, whereas the open circles to the estimation errors of the mean open-time. The results of the last of 800 iterations are illustrated.

estimated, probably because brief sojourns in other states, lasting one or two digitized points, were frequently missed. The overall estimation error was 14.2%. We have ascertained that the magnitude of errors systematically decreases as the signal amplitudes increase and vice versa.

(*e*) *Departure from the first-order Markov model assumption*

In the previous simulations, the signal sequences used were first-order Markovian. Here we show that the HMM processing scheme performs satisfactorily even when the signal sequence embedded in noise is not strictly first-order Markovian.

At two-state signal sequence was generated such that its closed-time interval distribution could be fitted with two exponential functions. The open-time distribution, on the other hand, was single exponential. A signal sequence, 47000 points in length and 0.1 pA in amplitude, was added to the noise ($\sigma_\omega = 0.1$ pA), and then MAP estimates were obtained by using the HMM processing method. Sample segments of the original signal, the signal in the noise and the corresponding estimates of the signal are displayed in figure 7*a–c*. From the estimated signal sequence, the open- and closed-time interval distributions were constructed (filled circles in figure 7*d*, *e*) and compared with those obtained from the original signal (open circles in figure 7*d*, *e*). Both open- and closed-time interval distributions obtained from the estimated signal sequence did not depart appreciably from the true distributions.

By using the same signal sequence of various amplitudes, we have also assessed the performance of the HMM processing method in characterizing signal statistics. In figure 7*f*, short segments of the noise and 5 fA, 10 fA, 15 fA and 25 fA signals are shown. The error magnitudes in estimating the amplitude and mean open-time of the signal are plotted against signal amplitude. When the signal amplitude was successively reduced from 200 fA to 5 fA, the estimation errors remained fairly constant at about 2 fA (filled circles in figure 7*g*). Thus, the fractional error increased as the signal amplitude decreased. Similarly, the estimates of the mean open-time (or a_{22}) increased progressively as the signal amplitude decreased (open circles in figure 7*g*). The estimated mean open-time was 3.39 ms (correct value = 3.33 ms) when the amplitude of the embedded signal was 200 fA, and the estimated value decreased steadily to 1.68 ms at 5 fA. Similar trends were observed for the estimates of the mean closed-time (not shown here).

From a number of simulations such as the one illustrated here, we conclude that the HMM processing scheme is relatively insensitive to the initial assumption that the underlying signal obeys the first-order Markov statistics. Thus, our method can be fruitfully applied to the analysis of typical channel data where the generating mechanism may consist of several closed or open conformational modes, of which some have short lives and some have very long lives.

4. SIGNALS BURIED IN NON-IDEAL NOISE

Biophysical data are frequently contaminated by, in addition to random noise, deterministic or non-random interferences. Among these are the periodic interferences from the electricity mains, which is very hard to eliminate completely from experimental environments, and drift of the baseline, which can be caused by slow changes in the junction potential between recording electrodes and the external ionic medium. Here we show that channel currents corrupted by periodic disturbances and baseline drift can be readily extracted and characterized.

(*a*) *Elimination of periodic interferences: low noise*

We consider periodic disturbances in time k of the form $\Sigma_{m=1}^{p} c_m \sin(\omega_m k + \phi_m)$ where the frequency components ω_m are known but the amplitudes c_m and phases ϕ_m are unknown. The periodic interference used for the following simulations consisted of the fundamental and its first harmonic. We note that the theory and the corresponding algorithm we have developed can handle, with a slight increase in computational steps, any number of higher harmonics.

The results of the simulation shown in figure 8 are derived from a segment of data in which the amplitudes of both sinusoidal interferences and channel currents are large compared with the background noise. A binary state Markov signal sequence, 0.2 pA in amplitude (figure 8*a*), was added to a noise trace that was contaminated by periodic disturbances. The added periodic interference (not shown) was of the form $c_1 \sin \omega_1 k + c_2 \sin \omega_2 k$, with $c_1 = c_2 = 0.2$ pA and the two frequency components corresponding to 50 and 100 Hz. From the observation sequence, a short segment of which is displayed in figure 8*b*, the HMM processing scheme extracted the maximum likelihood sequence of the periodic interference (figure 8*c*), the Markov signal sequence contaminated by the noise (figure 8*d*) and, finally, the maximum likelihood estimate of the signal sequence (figure 8*e*). The original periodic interference that was added was indistinguishable from the estimated sequence shown in figure 8*c*. The estimated signal sequence displayed in figure 8*e* faithfully mirrors the original signal sequence (figure 8*a*), with the exception that three brief events lasting one or two points (100 or 200 μs) were undetected. From the entire data segment analysed, we constructed the open-time (figure 8*f*) and closed-time (figure 8*g*) histograms of the original signals (open circles) and estimated signals (filled circles). The solid lines drawn through the data were calculated from the estimated transition matrix, using equation 4. The conspicuous estimation errors were the failure to detect the brief channel events.

(*b*) *Elimination of periodic interferences: high noise*

The usefulness of our processing technique can be more convincingly shown than with the previous case by analysing a data segment in which neither periodic interferences nor Markov signals are discernible by

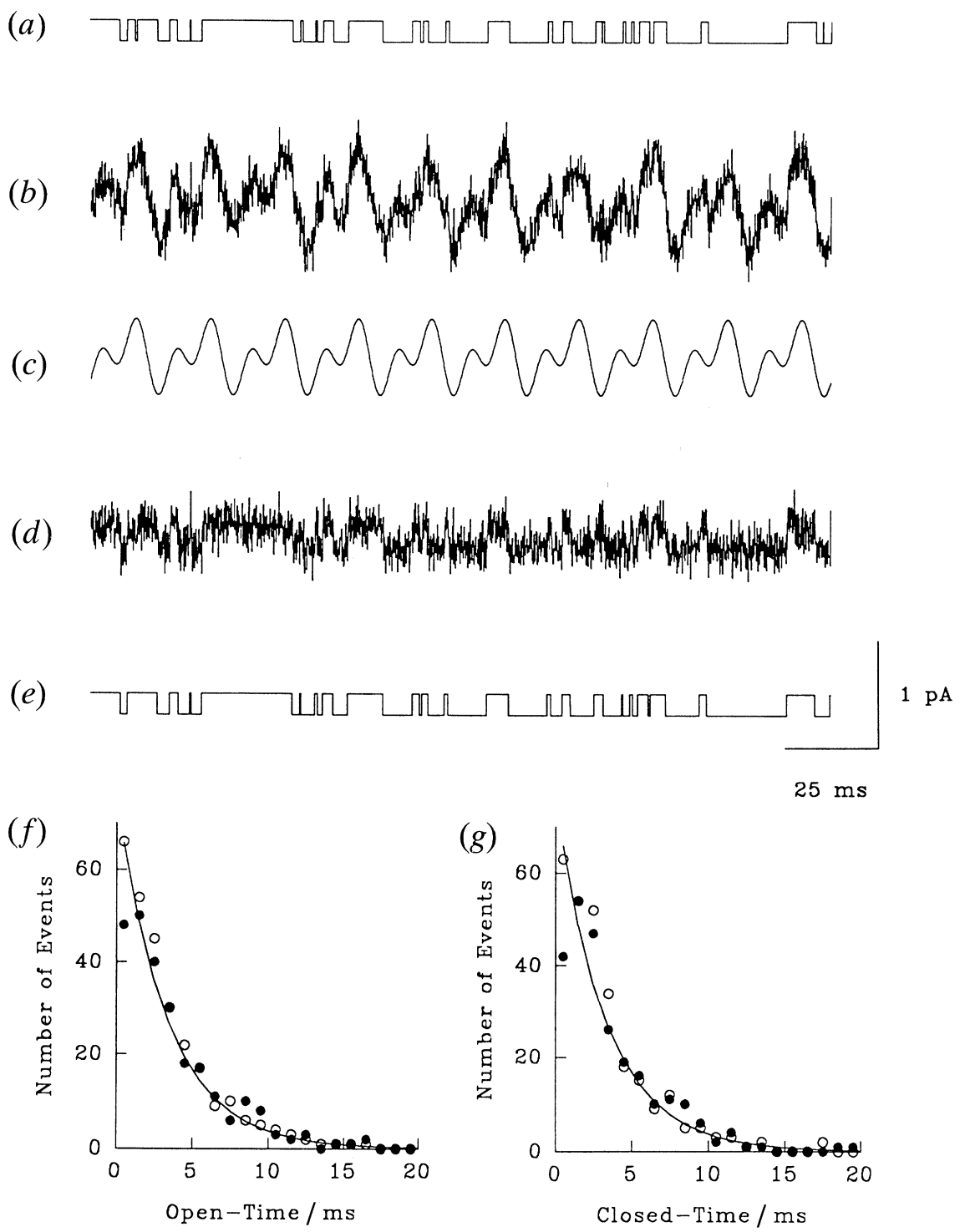

Figure 8. Characterization of a signal sequence buried in sinusoidal interference. A two-state Markov signal of amplitude 0.2 pA (*a*) and a sinusoidal wave composed of 50 and 100 Hz components (not shown) were added to noise to obtain the observation sequence (*b*). The amplitude of the two frequency components was 0.2 pA (phase = 0). The initial guesses used to process the data were: signal amplitudes, −0.1 pA and −0.3 pA; $a_{ii} = 0.9$; amplitudes of 50 and 100 Hz sinusoids, 0.15 pA, phases, 0.5 radians. The results obtained after 500 iterations are summarized in (*c–g*). The estimated periodic disturbance, shown in *c*, is indistinguishable from the original sinusoidal interference. The estimated parameters were 0.20 and 0.199 pA for the amplitudes of the two components, with their phase correctly given as zero radian. Also shown are the estimated observations sequence in the absence of the sinusoidal interference but in the presence of Gaussian noise (*d*) and the estimated signal sequence (*e*). The amplitudes of the signal were estimated to be −2.2 fA and −199.3 fA. Open-time (*f*) and closed-time (*g*) histograms were constructed from the estimated signal sequence (closed circles) and the original signal sequence (open circles). The solid lines drawn through the points are calculated from the estimated transition matrix.

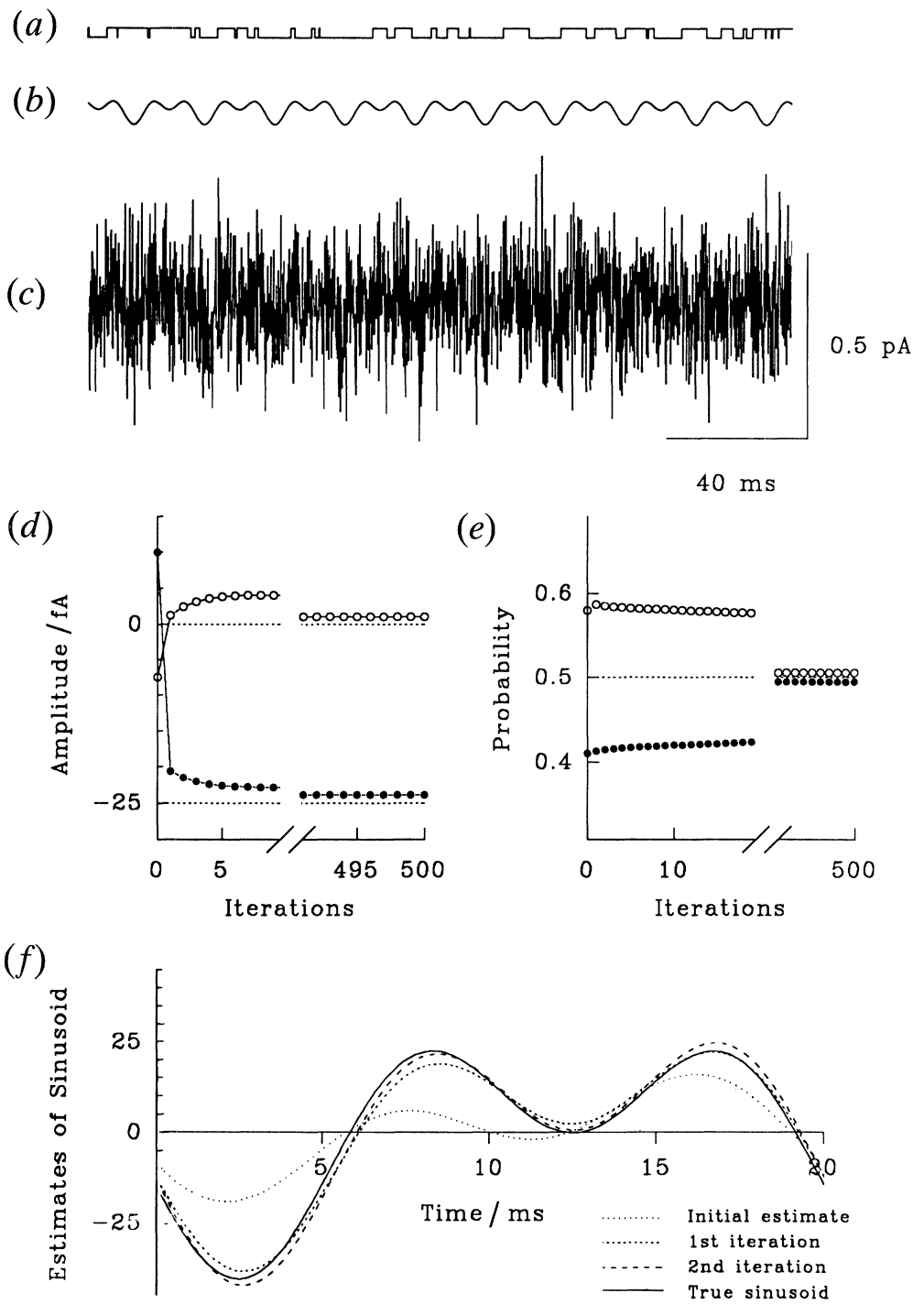

Figure 9. Characterization of a Markov signal sequence of small amplitude contaminated by the periodic interference. The Markov signal sequence (*a*) was generated by using the same parameters as in figure 7, except its amplitude was reduced to −25 fA. The periodic interference (*b*) added to the noise was composed of 50 Hz and 100 Hz, with the amplitude and phase of both components being 20 fA and $\pi/4$ radian. A segment containing the signal, noise and sinusoidal interference is shown in (*c*). The initial model parameters were: the amplitudes of the signal, +8 fA and −10 fA; the amplitudes and phases of the two sinusoidal components, 10 fA and 0.5 radian. The successive estimates of the signal amplitudes and dwell-time probabilities are shown in (*d*) and (*e*), with the correct values indicated by dashed horizontal lines. The estimates slowly converged to the correct values. The estimated parameters of the sinusoidal interference, in contrast, rapidly converged to the true values. In (*f*), the first two estimates and the true sinusoid, together with the initial guess, are plotted on an expanded timescale.

eye. The amplitudes of the signal sequence (figure 9*a*) and of the periodic interferences (figure 9*b*) added to the noise were $\frac{1}{4}$ of σ_ω. From the trace shown in figure 9*c*, the presence of a sinusoidal wave and discrete Markov signals is not immediately obvious. The successive estimates of the signal amplitude and the probabilities of being in the open and closed states are shown in figure 9*d*, *e*. The final estimates of the amplitudes of the two Markov states were 1.1 fA and −23.9 fA, compared with the true values of 0 and

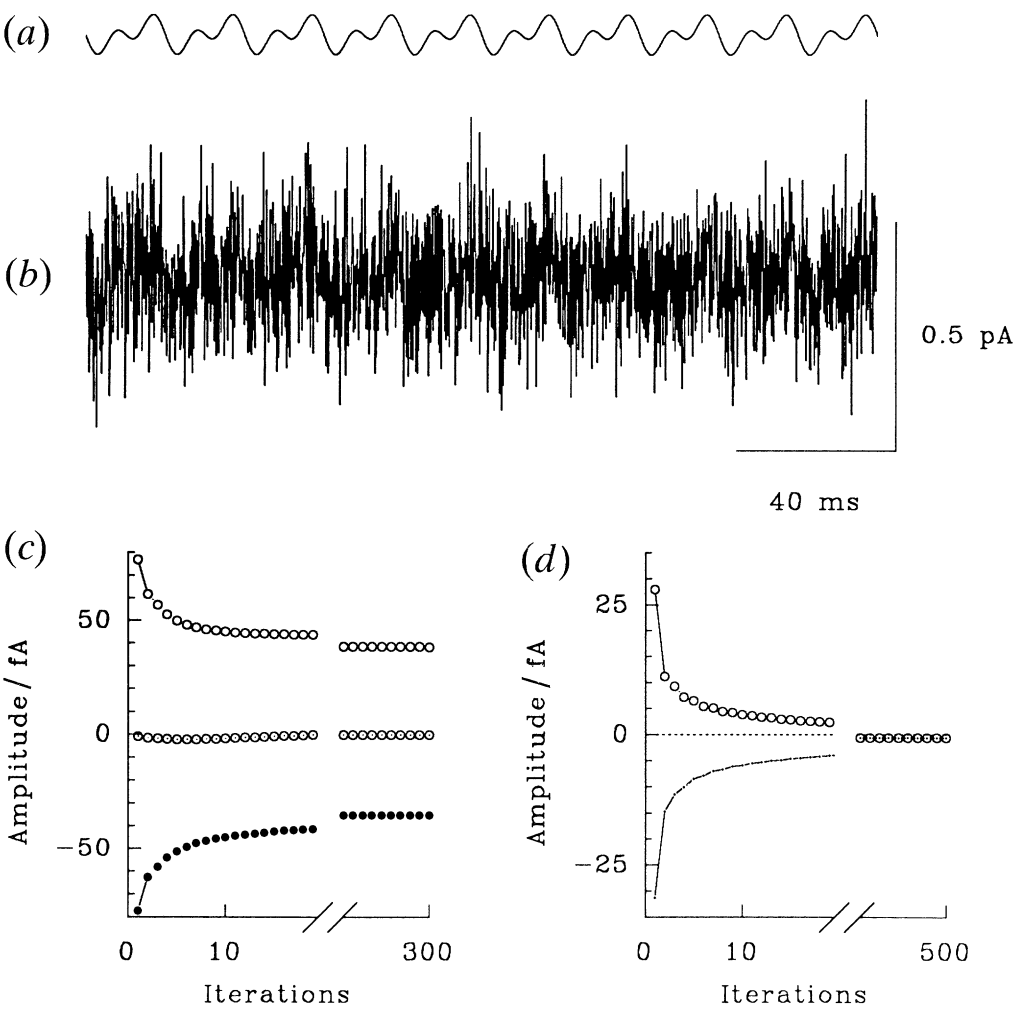

Figure 10. Discrimination between a signal sequence and sinusoidal interference. Sinusoidal interference composed of 50 and 100 Hz components (*a*) was added to noise to give the observation sequence (*b*). The amplitude of the two sinusoids, with no phase lag, was 25 fA. The observation sequence (*b*) was then analysed, erroneously assuming that it contained a Markov signal sequence. The estimated amplitudes at the first 20 and last 10 iterations are shown in (*c*). The observation sequence was re-analysed, allowing the presence of the periodic disturbance as well as the two-state Markov signal. The estimated amplitudes of the signal at the first 20 and the last 10 iterations are shown in (*d*). The two states were estimated to be identical (0.08 fA separation), indicating that the observation sequence contained no Markov signal.

25 fA. The probabilities of being in the closed and open states were estimated to be 0.495 and 0.505, respectively, the correct values being 0.5. The largest discrepancies between the true and estimated parameters were for the mean closed and open times. The estimated mean open and closed times, calculated from the estimated transition probabilities, were 2.89 ms and 2.58 ms, respectively, whereas the true means for both were 3.3 ms. In figure 9*f*, one period of the original sinusoid is compared with the estimates obtained at three successive iterations. By the third iteration, the estimated waveform was very close to the true one.

(*c*) *Detection of absence of Markov signals*

When the HMM processing scheme reveals the presence of small signals in noise-dominated data, it is imperative to eliminate the possibility that the detected signals stem from environmental interferences. In practice, we find it easy to distinguish between Markov signals and sinusoidal disturbances. By using the processing scheme, we illustrate here how a detected signal can be correctly attributed to the underlying sinusoidal wave that is buried in the noise.

A periodic wave composed of two sine waves, 50 Hz and 100 Hz, of equal amplitude (25 fA) was added to the noise. Sample segments of the periodic wave alone and the noise containing this wave are shown in figure 10*a*, *b*. The results of the analysis with the HMM processing scheme under the erroneous assumption that no periodic disturbances were present are shown in figure 10*c*. The amplitudes of three levels were identified to be −0.2 fA (baseline), +38.3 and −35.3 fA, with the probabilities of occupying these states being, respectively, 0.44, 0.26 and 0.30. The approximate symmetry of the signal amplitudes above and below the baseline renders it unlikely that such a signal sequence could be biological in origin. This assertion also can be readily confirmed by examining the power spectrum of the data. A spectrum obtained from the first 2000 points, not shown here, using the

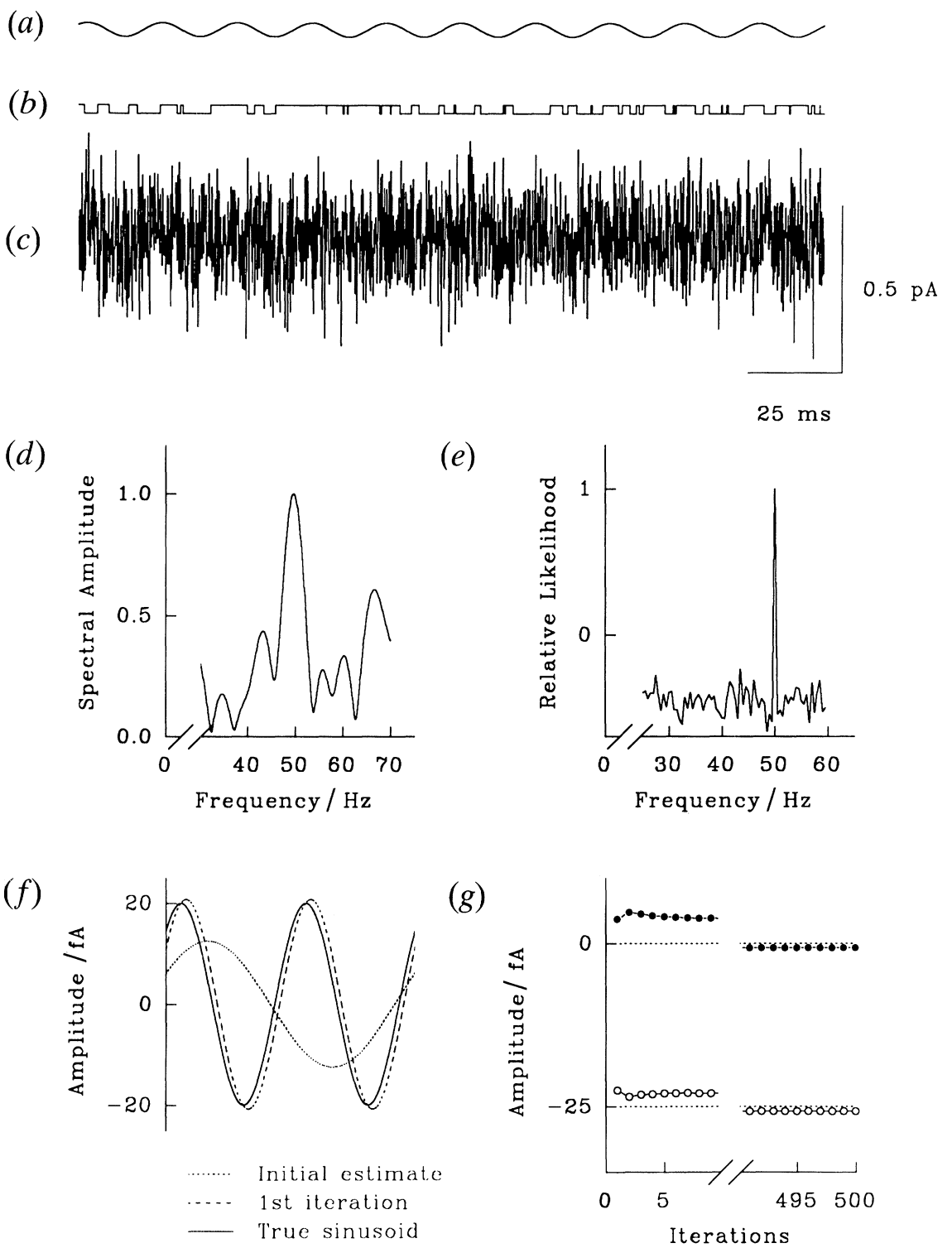

Figure 11. Estimation of the frequency of sinusoidal interference together with other relevant parameters. A 50 Hz sinusoidal wave with amplitude 20 fA and phase of $\pi/4$ rad (*a*) and a signal sequence with amplitude −25 fA (*b*) were added to noise to give the observation sequence (*c*). The amplitude spectrum, obtained by performing a zoom FFT at the resolution of 0.1 Hz on the first 2048 point of the data, gave a broad peak centred at 49.7 Hz (*d*), the magnitude error of which was unacceptably large (0.6%). The observation sequence was analysed with the processing scheme. The relative likelihood obtained after the first iteration is plotted against the frequency in (*e*). At this stage, the global maximum occurred at 49.94 Hz. The estimated frequency after 500 iterations was 49.99 Hz. The initial guess and the first estimate for the sinusoidal interference, together with the sinusoid which was initially embedded in the noise, are shown in (*f*) on an expanded time scale. The estimated amplitudes of the signal at the first and last ten iterations are shown in (*g*).

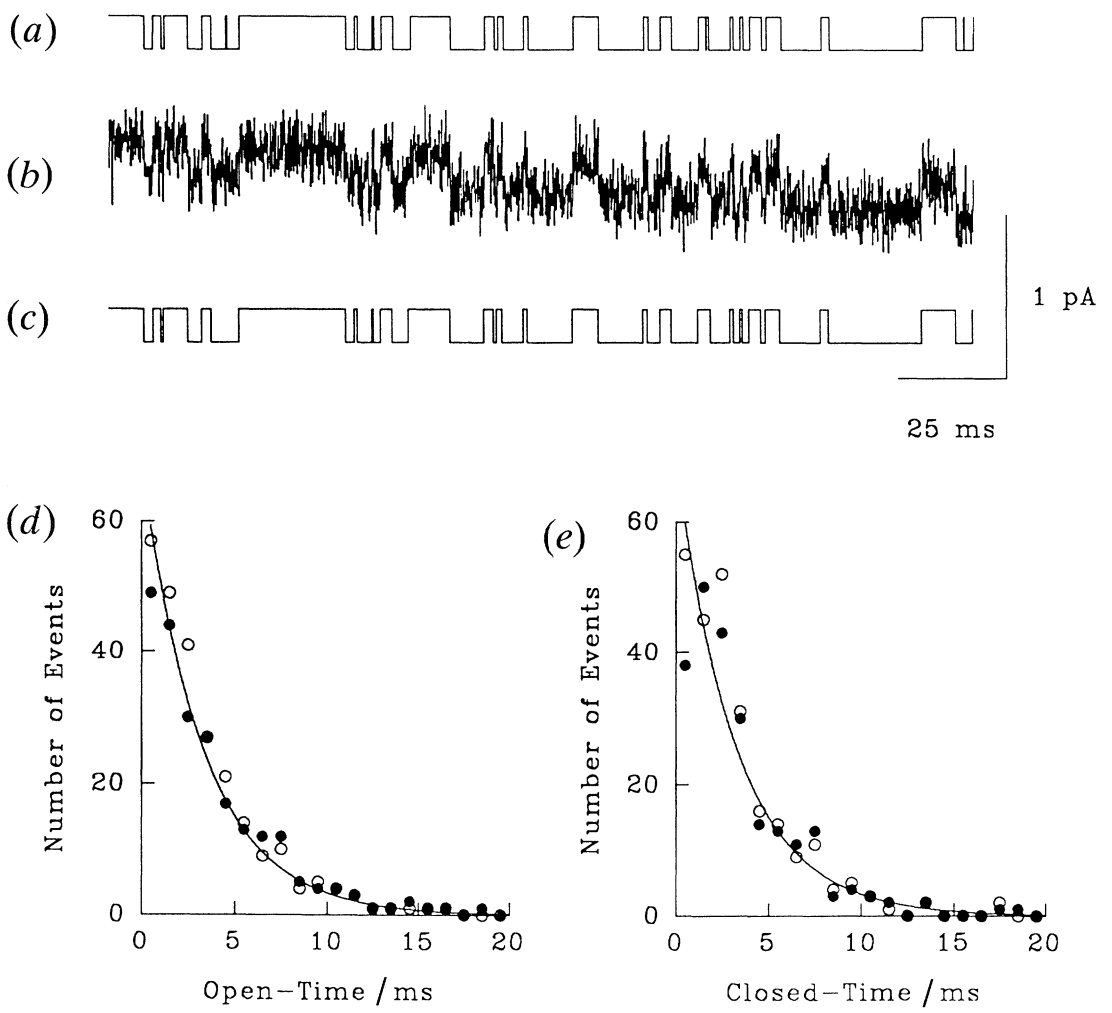

Figure 12. Extraction of a Markov signal in the presence of a steady baseline drift. A two-state Markov signal sequence with amplitude 0.2 pA (*a*) was added to Gaussian noise which drifted linearly downwards at a rate of 1.35 pA s^{-1}. From the observation sequence (*b*), the statistics of the signal, its maximum *a posteriori* sequence estimate (*c*), and the constant of the drift were estimated. The open-time and closed-time histograms of the original signal sequence (open circles) and estimated signal sequence (filled circles) are shown in (*d*) and (*e*). The solid lines drawn through the data points are calculated from the estimated transition probabilities.

Maximum Entropy Method, revealed two prominent peaks, one at 50 Hz and the other at 100 Hz, indicating that the data were contaminated by the periodic disturbances. The correct answers emerged when the analysis allowed the presence of the periodic disturbance (figure 10*d*). The estimated amplitude of the Markov signal was near zero (0.025 fA), showing that no signal was present. The amplitudes of the two sinusoids were estimated to be 26.6 fA and 25.6 fA, as compared with the true values of 25 fA. Their estimated phases were 0.032 and −0.037 radians, close to the true values of zero.

(*d*) *Re-estimation of the frequencies of the sinusoids*

Although the frequency of the electricity mains, which is the primary source of periodic disturbances in electrophysiological recordings, does not in general depart appreciably from 50 Hz, it is nevertheless desirable to devise a scheme whereby the exact interfering frequencies can be found adaptively. By using the formulae given in the Appendix (equations A 3–A 5), we have re-estimated the frequency of the periodic interference, in addition to the other relevant parameters of the Markovian signal.

Added to the noise were a 50 Hz sinusoidal wave of amplitude 20 fA and phase $\pi/4$ radian (figure 11*a*) and a binary Markov signal sequence, the two states being separated by 25 fA (figure 11*b*). When added to the noise, the presence of the periodic interference and the signal were not apparent (figure 11*c*). Assuming that the frequency, amplitude and phase of the periodic interference are unknown, along with the amplitude and characteristics of the Markov signal, we estimated these unknown parameters using the algorithm implemented according to equation A 5 given in the Appendix. In high noise, estimates of the frequency components of the periodic disturbance obtained by a zoom Fast Fourier Transform (FFT) on the data are not sufficiently accurate, as shown in figure 11*d*. A zoom FFT performed on the first 2048 point data at 0.1 Hz resolution shows a broad amplitude spectrum with a broad peak centred at 49.7 Hz. In contrast, the algorithm we implemented provided an unambiguous estimate of the frequency, amplitude and phase of the

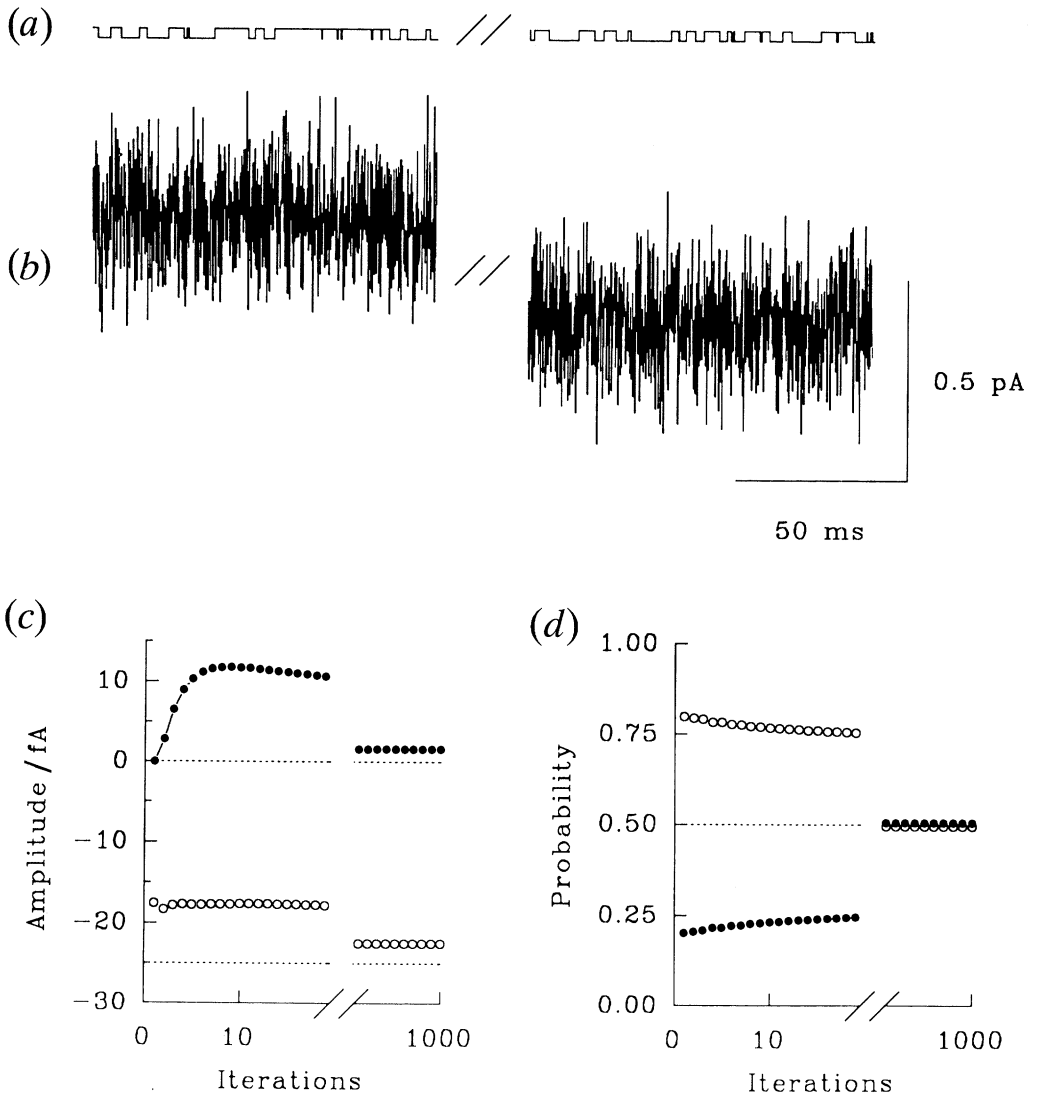

Figure 13. Characterization of small signals in the presence of baseline drift. A two-state Markov signal sequence with amplitude 25 fA (*a*) was added to noise which drifted linearly downward at a rate of 250 fA s^{-1}. Shown in (*b*) are two short segments of the observation sequence, the second trace being taken 1 s after the first. Estimates of the amplitude and probabilities of dwell-time at the first 20 and last 10 iterations are shown in (*c*) and (*d*). Dotted lines represent the true values. The DC shift in the results, we believe, is due to a small deviation from zero in the mean of the computer-generated noise from zero.

underlying periodic disturbance. The initial guess for the frequency was 25 Hz. The re-estimated frequency after the first iteration, as shown in figure 11*e*, was 49.94 Hz. Here, the ordinate is the normalized likelihood of the periodic disturbance at each discrete frequency (equation A 5). With each successive iteration, the estimated parameters of the periodic disturbance improved, slowly approaching the correct values. This is illustrated in figure 11*f*, where the first two cycles of the initial guess, the first re-estimate, together with the original periodic disturbance, are reproduced. The estimated amplitude, frequency and phase of the periodic disturbance after 500 iterations were, respectively, 21.29 (20) fA, 49.99 (50.0) Hz and 0.62 (0.785) radian. The numbers in the parenthesis are the true values. The estimated amplitudes of the signal sequence, shown in figure 11*g*, also approached the true values. The final estimates of the two state levels were −0.7 fA and −25.7 fA, as compared with the true values of 0 and −25 fA. The transition probability estimates were $a_{11} = 0.968$ and $a_{22} = 0.969$. The true probability for a_{ii} used to generate the original signal was 0.97.

(*e*) *Adjustment of baseline drift: low noise*

The level of the baseline in an experimental situation frequently drifts either slowly away from the starting point or abruptly steps to a new level and then returns to the original level. The second type is difficult to deal with, as it is not always possible to ascertain whether such a stepwise change represents an artefactual drift or a channel opening to one of its subconductance levels. A slow, erratic and continuous drift of the baseline, on the other hand, is unambiguous to the experimenter and can be easily eliminated with our processing method. We have incorporated the scheme whereby such a drift, if it exists, is adaptively corrected and provides the estimates of the channel statistics taking the unsteady baseline into account (equation

A 2). In the following two examples, we assumed for simplicity that the baseline drifts downwards linearly in time. We note, however, the processing technique we devised is for the general form of drift that can be represented by a polynomial function of time.

To a binary Markov signal sequence contained in the noise, a downward drift of 1.35 pA s^{-1} was introduced. The amplitude of the Markov signal, 0.2 pA, was twice that of σ_ω. Sample segments of the signal and noisy signal are exhibited in figure 12*a*, *b*. By using the HMM processing scheme, we estimated the rate of drift, as well as all the relevant statistics of the signal embedded in the steadily drifting noise. The estimated drift rate, after 1000 iterations, was 1.3496 pA s^{-1}. The two Markov states were estimated to be at +0.001 and −0.1995 pA, compared with the true values of 0 and −0.2 pA. The estimated probabilities of being in the open and closed states were 0.51 and 0.49, compared to the correct values of 0.5. The magnitudes of these estimation errors are within the acceptable range for most biophysical applications. A segment of the estimated signal sequence is shown in figure 12*c*. In the 2000 point segment illustrated, the processing scheme failed to detect three events, all brief openings or closings lasting 100 μs. In figure 12*d*, *e*, we constructed the open- and closed-time histograms of the original signal sequence (open circles) and estimated signal sequence (filled circles). The solid lines fitted through the data points are calculated from the estimated transition matrix, according to equation 4.

(*f*) *Adjustment of baseline drift: high noise*

A binary Markov signal sequence of 25 fA in amplitude (figure 13*a*) was added to the noise, which was drifting linearly downward at the rate of 250 fA s^{-1}. The initial 1000 point record and the same length of the segment taken 1 s later are shown in figure 13 *b*. With the processing scheme, we iteratively estimated the amplitude, transition probability and relative occupancy probability of the Markov signal as well as the rate of drift. In figure 13*c*, the estimates of the signal amplitude are plotted against successive iterations. The final estimates of the signal levels were +1.63 fA and −22.5 fA, compared with the true values of 0 and −25 fA. The probabilities of being in each of the two states were estimated to be 0.495 and 0.505, giving 1% errors. The estimated transition probabilities, a_{11} and a_{22}, were both 0.967, the correct values being 0.97. Finally, the rate of drift was estimated to be 251.4 fA s^{-1}.

5. DECOMPOSITION OF TWO INDEPENDENT CHANNELS

One of the problems often encountered in single channel recordings is that more than one channel is contained in an isolated patch of the membrane. Moreover, when a current trace shows multiple levels, it is sometimes difficult to determine whether different levels represent the subconductance states of a single channel or independent openings and closings of two or more single channels contained in the patch. With these problems in mind, we have devised a processing scheme with which we can decompose a record containing two or more single channels which open and close independently of each other.

Simulations, not presented here, showed that if the amplitudes of the two or more independent Markov chains were identical, the estimates of the individual chain was not as accurate as when their amplitudes were different. For example, let us consider the case when two Markov chains are present. Suppose both Markov chains have state level 0 and 1. Then the sum of the two chains is a process with three levels at 0, 1 and 2. In such a case, the state level 1 is ambiguous because it could have occurred with the first chain at level 1 and the second at 0 or vice versa. It is this ambiguity together with the finite data length which degrades the performance. However, if the transition probabilities of the two chains are significantly different, then the effect of the state ambiguity is diminished and the estimates of the statistics of the chain improve.

(*a*) *Two Markov chains in ideal noise*

A signal sequence, a 2000 point segment of which is shown in figure 14*a*, was generated by adding two Markov signal sequences of different amplitudes and different transition matrices. This summed signal sequence was then embedded in noise (figure 14*b*). From the noisy record it is not immediately obvious whether the underlying signals represent the algebraic sums of two independent single channels or single channel events showing multiple conductance levels. Assuming erroneously, for the first instance, that underlying the signal sequence was an N-state, first-order Markov process, we obtained its characteristics by using the standard processing method. The results of the analysis correctly revealed that there were four current levels, in fA at 0.1 (0.0), 199.9 (200), 297.9 (300) and 500.1 (500) with the relative occupancy probabilities of 0.625, 0.298, 0.063 and 0.024, respectively. The numbers in the parentheses are the true levels. A cursory inspection of the 4×4 transition matrix made it apparent that the record contained the activities of two independent channels. The first and fourth row of the estimated transition matrix read: (0.97, 0.02, 0.01, 0) and (0, 0.10, 0.05, 0.85), showing that the transitions from the baseline to the highest state or vice versa did not occur. In other words, the probability of two independent channels opening or closing precisely at the same instant was zero. Similarly, both a_{23} and a_{32} were close to zero.

Once it is known that two independent Markov signal sequences are contained in the record, the summed signal sequence can be readily decomposed to individual signal sequences, and the characteristics of each sequence can be obtained. Clearly, in such a case the combined transition matrix $\boldsymbol{A}_T$, the subscript T denoting tensor, will be a tensor product of two matrices, $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$. The processing technique we have devised provides the estimates of $\boldsymbol{A}_T$ and two Markov signal sequences. The original signal and estimated signal sequences of the first Markov process are shown

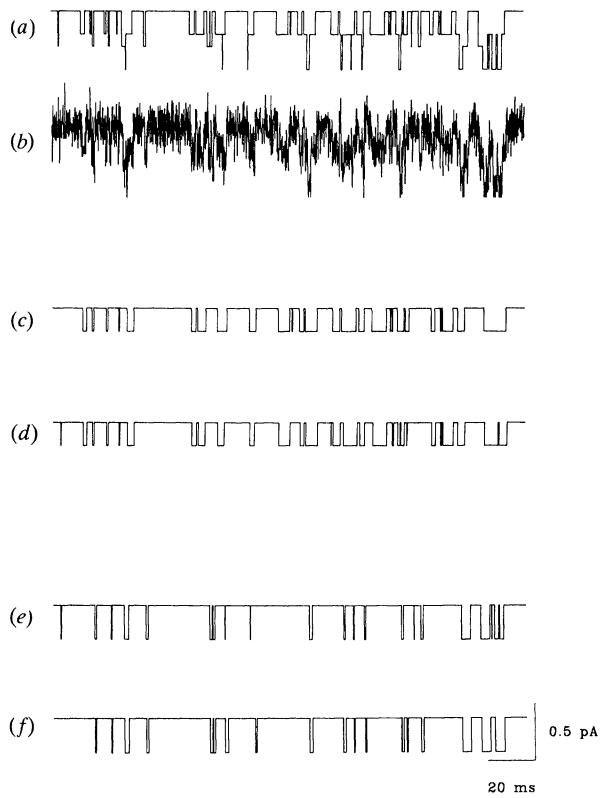

Figure 14. Decomposition of two independent Markov signals. A segment of the signal sequence (*a*), obtained by adding two independent Markov signals, was combined with noise. The signals were generated according to the transition matrices, $a_{11} = 0.98$ and $a_{22} = 0.95$ for the first Markov process, and $a_{11} = 0.99$ and $a_{22} = 0.90$ for the second. Their respective amplitudes were 0.2 pA and 0.3 pA. From the statistics obtained for the combined signals, it could be ascertained that the signal sequence embedded in the noise represented the sum of two independent Markov processes (see text). By using vector HMM processing, the signal sequence was decomposed into its constituents. Displayed are short segments of the original signal sequence of the first process (*c*) with its estimated sequence (*d*), and the original signal sequence of the second process (*e*) with its estimated sequence (*f*).

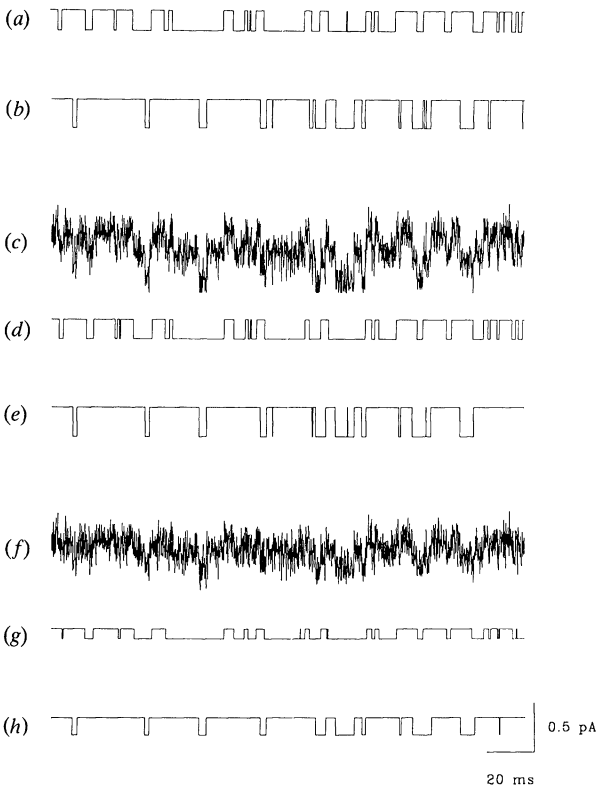

Figure 15. Increasing estimation errors with decreasing amplitudes of two independent Markov processes. Two signal sequences were generated according to the transition probabilities, $a_{11} = 0.98$ and $a_{22} = 0.98$ for the first sequence (*a*) and $a_{11} = 0.99$ and $a_{22} = 0.95$ for the second sequence (*b*). The amplitudes of the signals were systematically reduced, and after adding them to noise, the observation sequence was analysed as detailed in figure 14. The results of two such simulations are illustrated. The trace shown in (*c*) contains the two-state Markov processes with amplitudes of 0.15 pA and 0.225 pA. The decomposed signal sequences of the two processes are shown in (*d*) and (*e*), corresponding to the original signal sequences shown in (*a*) and (*b*). In trace (*f*), the same signal sequences were added to noise, except that their amplitudes were now reduced to 0.1 pA and 0.175 pA. The decomposed signal sequences are shown in (*g*) and (*h*). Further reductions in signal amplitudes caused progressively larger estimation errors.

in figure 14*c*, *d*, and the corresponding segments of the second Markov process are shown in figure 14*e*, *f*. The estimation errors occurred, as expected, predominantly with brief events, which were either missed or incorrectly assigned. The estimated $\boldsymbol{A}_T$ was decomposed into two matrices representing the transition probabilities of the first and second signal sequences:

$$\boldsymbol{A}_T = \begin{pmatrix} 0.968 & 0.032 \\ 0.053 & 0.947 \end{pmatrix} \otimes \begin{pmatrix} 0.987 & 0.013 \\ 0.103 & 0.897 \end{pmatrix}.$$

The true matrices used to generate the original signals were: $a_{11} = 0.98$, $a_{22} = 0.95$ for the first signal sequence, and $a_{11} = 0.99$, $a_{22} = 0.90$ for the second signal sequence. (Note: the decomposition of a tensor product into its constituent matrices is in general not unique but the solution with the restriction $\sum_{j=1}^{N} a_{ij} = 1$ is unique.)

The results illustrated in figure 15 show that the channel currents emanating from two single channels can be decomposed with an acceptable degree of accuracy even when their amplitudes are small relative to the noise. Two signal sequences, shown in figure 15*a*, *b* were first added and then embedded in the noise. By using the same procedures as in figure 14, the summed signals were extracted from the noise and decomposed into two individual sequences. The amplitudes of the two signal sequences shown in figure 15*c* were 0.15 pA and 0.225 pA. The estimated signal sequences are shown in figure 15*d*, *e*. When the signal amplitudes were further reduced to 0.1 pA and 0.175 pA, the estimated signal sequences (figure 15*g*, *h*) from the noisy data (figure 15*f*) contained significantly more errors than the previous example.

(*b*) *Two Markov chains in non-ideal noise*

To mimic real experimental data, we have generated a noisy record, albeit somewhat exaggerated, that contained a signal sequence which was not first-order

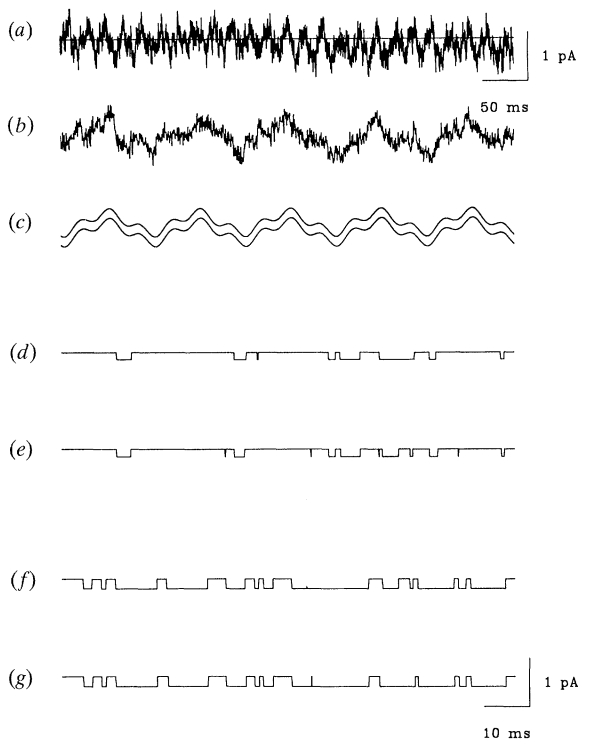

Figure 16. Two independent Markov chains buried in non-ideal noise. Two Markov processes were added and embedded in the noise. Then, a harmonic wave of the form $p_k(\Theta) = c_1 \sin(\omega_1 k + \pi/2) + c_2 \sin(\omega_2 k + \pi/4)$, with $c_1 = 0.2$ pA, $c_2 = 0.1$ pA, ω_1 and ω_2 corresponding respectively to 50 Hz and 150 Hz, and a linear drift of 0.5 pA s^{-1} were added. The amplitudes of the first and second Markov processes were, respectively, 0.15 pA and 0.2 pA. The first process, a first-order Markov, was generated with the transition probability matrix of $a_{11} = 0.99$ and $a_{22} = 0.95$, whereas the second process, not first-order Markovian, was the same signal sequence used for figure 7. The first 500 ms and 50 ms of the record are shown in (*a*) and (*b*). In (*c*), segments of the original harmonic wave (upper trace) and the estimated wave (lower trace) are displayed. After eliminating the harmonic wave and baseline drift, the signal contained in the noise was decomposed into two constituent processes. Segments of the first (*d*) and second (*f*) signal sequences are compared with the corresponding segments of the estimated sequences (*e* and *g*).

Markovian, and a smaller background signal sequence which opened and closed independently of the first chain. The simulated experimental record was heavily contaminated by interferences from the power line, composed of 50 Hz and 150 Hz, with different amplitudes and phases. In addition, the baseline was rapidly drifting downwards.

In figure 16*a*, *b*, the first 500 ms and 50 ms of the record are shown. The task of our fully implemented HMM processing method was to parcel out the interfering deterministic components, characterize the combined signal sequence, and then decompose it into two independent chains. The original sinusoidal wave added to the record is shown in figure 16*c*, together with the estimated wave directly below it. The two traces are indistinguishable when superimposed on the timescale used. Similarly, the rate of baseline drift was correctly estimated (not shown here). The processing method identified that the signal assumed, in addition to the baseline, three discrete amplitudes at 0.15 pA, 0.2 pA and 0.35 pA. But the transition from the baseline to the 0.35 pA level or vice versa did not occur, indicating that the signal sequence originated from the sum of two independent chains. The HMM processing scheme then decomposed the combined signal sequence into two independent chains. The sample segments illustrated in figure 16*d–g* are the original (*d* and *f*) and estimated (*e* and *g*) signal sequences of the first and second chain.

Table 1. *Estimates of the parameters of the deterministic interferences and of the statistics of the embedded Markov chains*

	true values	estimated values
sinusoidal components		
amplitude of 50 Hz/fA	200	201.4
phase of 50 Hz/deg	90	88
amplitude of 150 Hz/fA	100	100.2
phase of 150 Hz/deg	45	40
rate of drift/(fA s^{-1})	500	497.3
first Markov chain		
amplitude/fA	150	155.3
mean open-time/ms	2	2.37
second Markov chain		
amplitude/fA	200	201.7
mean open-time/ms	3.33	3.08

The estimates of the parameters of the deterministic interferences and of the statistics of the embedded Markov chains were acceptably accurate. These are summarized in table 1.

Faced with a set of imperfectly determined measurements, such as the one shown in figure 16*a*, the experimenter may be tempted to discard the data and embark upon a costly and time-consuming process of repeating the measurements. With the advent of modern digital signal processing techniques, which unlike conventional filters utilize all prior knowledge about the signal and unwanted random and patterned disturbances, useful information can be gleaned from such a real world process.

6. DISCUSSION

(*a*) *Hidden Markov Models and Expectation Maximization algorithm*

The HMM signal processing techniques we have studied and tested extensively can be satisfactorily applied, but their application is not limited to, in extracting and characterizing small transmembrane channel currents buried in background noise. The currents recorded during an experimental situation are not only corrupted by unavoidable random noise but also they are frequently contaminated by periodic disturbances originating from the electricity mains, composed of a fundamental 50 Hz frequency and odd harmonics, as well as baseline drift. The processing methods described in this paper yield estimates of these deterministic interferences, as well as all the relevant statistics of the underlying signal. Moreover, we have described a method for decomposing two or more independent Markov chains embedded in the noise.

When the open state of a channel exhibits multiple conductance levels, it is important to determine unambiguously whether the signal sequence represents an N-state Markov process or an algebraic sum of two or more independent Markov processes.

In devising the signal processing techniques, we have first formulated the problems in the framework of an HMM, and then applied the EM algorithm (Dempster *et al.* 1977; Titterington *et al.* 1985) to obtain the maximum likelihood estimates. We note here that there are alternative numerical methods for calculating the maximum likelihood estimates. One approach we have considered, and rejected for the reasons given below, is the Newton–Raphson algorithm which when it converges does so quadratically and thus rapidly. The EM algorithm, on the other hand, converges linearly, and so convergence can be very slow. However, with the Newton–Raphson algorithm, the computational steps involved tend to be complicated and the memory requirements to obtain the estimates are large, especially since the Hessian matrix needs to be inverted. Moreover, successive iterations with the Newton–Raphson algorithm do not necessarily improve the likelihood function. In contrast, the EM algorithm is simple to implement and satisfies the appealing property that the likelihood function is always improved after each iteration.

We have extensively tested the reliability of the algorithms we implemented in characterizing Markov signals buried in the noise. With ideal noise, which is white, Gaussian and contains no extraneous deterministic inferences, it was possible to characterize a Markov process whose levels were separated by 1/20 to 1/10 of the standard deviation of the noise (figures 4 and 5). The standard deviation of the noise from a patch-clamp amplifier with the tip of an electrode pipette tightly sealed with a membrane patch, when filtered at 2 kHz, is about 0.1 pA (0.2 pA when filtered at 5 kHz). Under these conditions, channel currents whose amplitudes are as low as 5 to 10 fA can be adequately characterized.

Unless there is an unambiguous method of distinguishing the recording artefacts from the real signal, the processing scheme we detailed in §3 will have limited biological application. The real electrophysiological data invariably contain, in addition to biological signals and amplifier noise, other recording artefacts, the most prominent of which is the periodic interference from the power line. This sinusoidal wave consists of, for obvious reasons, the fundamental and its odd harmonics, predominantly 50, 150 and 250 Hz. Notch filtering or a generalized notch filtering approach (Paiss 1990) has a considerable transient response and thus obscures and distorts the embedded Markov signal. The processing schemes we have formulated for eliminating the sinusoidal interference, as well as the baseline drift, fully exploit all the *a priori* information available: the nature of the deterministic disturbances, the Markovian characteristics of the signal and the presence of white, Gaussian noise. In this sense, the processing can be said to be optimal in that maximum likelihood estimates are obtained. The reliability of the processing schemes in characterizing the parameters of the deterministic interferences, together with those of the embedded Markov chain, is demonstrated in §§4 and 5. In practice, when reasonable care is taken to minimize interference from the mains, the amplitudes of the residual sinusoids present in the records are of the order of 20–50 fA for 50 Hz, 150 Hz and 250 Hz.

(*b*) *Signal models and underlying assumptions*

The processing methods we have detailed are based on two key assumptions. The noise corrupting the signal is Gaussian and white (memoryless), and the underlying signal is a first-order, finite-state, discrete-time, Markov process. The violation of the whiteness of the noise assumption severely degrades the performance of the processing scheme, whereas the departure from the first-order Markov assumption does not affect the estimates of the signal statistics appreciably. Thus, even with a signal sequence that is second- or higher-order Markovian, or a periodic step change occurring at a fixed interval, the extracted signal sequence and the estimates of its associated statistics are acceptably accurate. In contrast, in using this or any processing methods based on the HMM techniques and the EM algorithm, care must be taken to ensure that the noise spectrum is flat up to the Nyquist frequency. For further discussion on the subject, see Chung *et al.* (1990).

In the same theoretical framework of HMM techniques and the EM algorithm, the signal model can be further modified and extended, so making the signal processing schemes far more versatile than those described here. Instead of assuming that channel currents are generated by a first-order Markov process, we can represent the signal sequence as m-ary Markovian. The most general form of the signal model is one with time-varying transition probabilities, known also as discrete-time, semi-Markov processes. Here, the transition probability, instead of being constant, is an unknown function of the time after the process made a transition to or from a given state (time to the last transition). The fractal model, postulated by Liebovitch and colleagues (1987, 1989) is a special case of this generalized representation. A brief mathematical description of this extension is described elsewhere (Krishnamurthy *et al.* 1991*a*). Also, a Markov process, after entering one of its states, can be allowed to decay back in time exponentially or otherwise to the original state. Such a signal process was formulated as an augmented homogeneous HMM problem and a scheme for estimating this stochastic process, when its realizations is hidden in noise, has been devised (Krishnamurthy & Moore 1991). In this context, we note that techniques for estimating filtered Markov processes with additive noise are presented in Goutsias & Mendel (1988).

(*c*) *Computational and memory requirements*

The drastic increase in the signal-to-noise ratio attained with the HMM processing scheme must be weighed against the computational cost. In the

forward–backward scheme, the number of computational steps involved is of the order of $N^2 T$, where N is the number of Markov states and T is the number of data points, while the memory requirements are $O(NT)$. Typically, we analyse about a 100000 point record using less than ten allowed states. Although a large number of computational steps are involved, the processing cost and time, we feel, are negligible compared with those expended for data acquisition. Because a modern workstation computer can perform about 25 million instructions per second, the real time involved in processing such a record, once the codes are optimized, will be of the order of minutes. It may be possible to increase the speed of processing using similar techniques to those suggested by Peters & Walker (1978). They propose a method of improving the convergence of EM when the embedded data is from an 'independently and identically distributed' process, that is, the process when $a_{1j} = a_{2j} = \ldots = a_{Nj}$. It remains to be investigated if it is possible to improve convergence similarly when the embedded process is Markov.

The re-estimation procedures developed in this paper are off-line. The entire sequence of observations is required for off-line processing and also the memory requirement $O(NT)$ can be large. In addition, the estimates of the transition probabilities and signal levels are updated only at the end of the iteration. It is of interest to develop on-line processing schemes which update the transition probabilities and state levels at each time instant when a new observation is available. Such procedures could be constructed to significantly reduce memory requirements. Titterington (1984) has developed on-line techniques when the embedded data is from an 'independently and identically' distributed process. We are currently developing on-line schemes for HMM processing. Such on-line schemes could also be used for eliminating periodic interferences with slowly varying frequencies, amplitudes and phases.

(*d*) *Concluding remarks*

One important use of our processing methods which we have alluded to but not specifically addressed is the evaluation of signal models (Rabiner & Juang 1986; Rabiner 1989). Conventionally, the distribution of open-time or closed-time histograms, accumulated from a relatively long segment of data, has been used to discriminate between different models (see, for example, Sansom *et al.* (1989)), but associated with this method are several obvious disadvantages (Liebovitch 1989; McManus *et al.* 1989). Given the observation, what is the most likely signal model? The same segment of the data can be analysed by using different signal models, and comparing the likelihood functions, as given in equation 3. Theoretically, the best model, or the model that is most consistent with the observation, is the one which gives the highest value of the likelihood function. Such a mathematical tool for discriminating unambiguously a class of plausible models from implausible ones may prove to be useful for the understanding of the molecular mechanisms underlying channel openings. If the analysis of channel currents is motivated by a model that is inadequate, the effort expended in deriving its parameters or kinetic constants may turn out to be futile. Moreover, characterization of channel currents which are at least an order of magnitude smaller in amplitude and more brief in duration than those amenable for analysis with conventional methods may ultimately provide a new insight into the dynamics of protein macromolecules forming ionic gates in living membranes.

This work was in part supported by a grant from the National Health and Medical Research Council of Australia. Throughout the course of this study, Mrs Jennifer Edwards provided excellent technical assistance, for which we are grateful.

APPENDIX

Re-estimation formulae for eliminating deterministic interferences

We briefly summarize some of the results derived in a companion paper by Krishnamurthy *et al.* (1991*b*). We stress that the re-estimation equations presented here are not strictly based on the EM algorithm. Consider the observation sequence Y_T which contains an N-state Markov signal sequence, the periodic disturbance of the form $\sum_{m=1}^{p} c_m \sin(\omega_m k + \phi_m)$, or a drift in the states of the Markov process in the form of the polynomial $\sum_{n=1}^{p} d_n k^n$ and additive white Gaussian noise. Unknown are the amplitudes q_i of the Markov states and their transition probabilities a_{ij}, the amplitudes c_m an the phases ϕ_m and the constant d_n of the drift. The problem is to obtain the maximum likelihood estimates of these unknown parameters. The solution of this problem involves the EM algorithm, an iterative algorithm consisting of the Expectation step and the Maximization step (Dempster *et al.* 1977).

Let $p_k(\boldsymbol{\Theta}), \boldsymbol{\Theta} \in \boldsymbol{R}^R$, with unknown parameter vector $\boldsymbol{\Theta} = (\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_R)$, denote a deterministic disturbance, either a periodic or polynomial drift disturbance or both. The expectation of the log of the likelihood function $\zeta(\lambda, \hat{\lambda})$ of the 'fully categorized data' (Titterington *et al.* 1985) may be expressed as:

$$\zeta(\lambda, \hat{\lambda}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{T-1} \xi_k(i,j) \log \hat{a}_{ij} + \sum_{i=1}^{N} \gamma_k(i) \log \hat{\pi}_i + \zeta_2, \tag{A 1}$$

where

$$\zeta_2 = \sum_{k=1}^{T-1} \sum_{i=1}^{N} \gamma_k(i) \log \frac{1}{\sqrt{2\pi}\,\hat{\sigma}_\omega} \times \exp\left(-\frac{(y_k - (\hat{q}_i + p_k(\hat{\boldsymbol{\Theta}})))^2}{2\hat{\sigma}_\omega^2} \right).$$

The Maximization step involves finding $\hat{\lambda}$ to maximize $\zeta(\lambda, \hat{\lambda})$. Now we consider polynomial (drift) and periodic (sinusoidal) disturbances. Solving for $\boldsymbol{\Theta}_n$ in $\partial \zeta_2 / \partial \hat{d}_n = 0$, for $\boldsymbol{\Theta} = (d_1, \ldots, d_p)$ and $p_k(\boldsymbol{\Theta}) = \sum_{n=1}^{p} d_n k^n$, it can be shown that the re-estimation formulae for the polynomial drift constants are the solutions of the p equations which are linear in d_n

$$\sum_{k=1}^{T} \sum_{i=1}^{N} \gamma_k(i) \, (y_k - \hat{q}_i) \, k^n = \sum_{n=1}^{p} p_k(\hat{\boldsymbol{\Theta}}) \, k^n, \quad n = 1, \ldots, p. \tag{A 2}$$

For $\Theta = (c_1, \ldots, c_p, \phi_1, \ldots, \phi_p)$ and $p_k(\Theta) = \sum_{m=1}^{p} c_m \sin(\omega_m k + \phi_m)$, the amplitude estimates are obtained as:

$$\hat{c}_m = \frac{\sum_{k=1}^{T}\sum_{i=1}^{N}\gamma_k(i)\left(y_k - \hat{q}_i - \sum_{\substack{m=1\\ m\neq n}}^{p}\hat{c}_m \sin(\omega_m k + \hat{\phi}_m)\right)\sin(\omega_n k + \hat{\phi}_n)}{\sum_{k=1}^{T}\sin^2(\omega_n k + \hat{\phi}_n)}. \quad \text{(A 3)}$$

Similarly, the phase estimate ϕ_m is obtained by solving:

$$M_1 \cos\hat{\phi}_m + M_2 \sin\hat{\phi}_m = N_1 \cos 2\hat{\phi}_m + N_2 \sin 2\hat{\phi}_m, \quad \text{(A 4)}$$

where

$$M_1 = \sum_{k=1}^{T}\sum_{n=1}^{N}\gamma_k(i) \times\left(y_k - \hat{q}_i - \sum_{m\neq n}^{p}\hat{c}_n \sin(\omega_n k + \hat{\phi}_n)\right)\cos(\omega_m k),$$

$$M_2 = -\sum_{k=1}^{T}\sum_{n=1}^{N}\gamma_k(i) \times\left(y_k - \hat{q}_i - \sum_{m\neq n}^{p}\hat{c}_n \sin(\omega_n k + \hat{\phi}_n)\right)\sin(\omega_m k),$$

$$N_1 = \frac{\hat{c}_m}{2}\sum_{k=1}^{T}\sin 2\omega_m k; \quad N_2 = \frac{\hat{c}_m}{2}\sum_{k=1}^{T}\cos 2\omega_m k.$$

In the above results, it was assumed that the frequencies of the sinusoids are precisely known. If the frequency components are also to be estimated, the parameter vector to be estimated is $\Theta = (c_1, \ldots, c_p, \phi_1, \ldots, \phi_p, \omega_1, \ldots, \omega_p)$. Again, the maximum likelihood estimate ω of $\omega = (\omega_1, \ldots, \omega_p)$ is obtained by solving:

$$f_i(\hat{\omega})) \overset{\Delta}{=} \sum_{k=1}^{T} h_k \cos(\hat{\omega}_m k + \hat{\phi}_m) - \hat{c}_m k \sin(2(\hat{\omega}_m k + \hat{\phi}_m)) - k\sum_{\substack{n=1\\ n\neq m}}^{p}\hat{c}_n \sin(\hat{\omega}_n k + \hat{\phi}_n)\cos(\hat{\omega}_m k + \hat{\phi}_m) = 0, \quad \text{(A 5)}$$

where

$$h_k = k\left(y_k - \sum_{i=1}^{N}\gamma_k(i)\,\hat{q}_i\right) \quad \text{and} \quad \hat{\omega} = (\hat{\omega}_1, \ldots, \hat{\omega}_p)^T.$$

Equation (A 5) can be solved numerically by using, for example, the Newton–Raphson method.

REFERENCES

Ascher, P. & Nowak, L. 1988 Quisqualate- and kainate-activated channels in mouse central neurones in culture. *J. Physiol., Lond.* **399**, 227–245.

Baum, L. E. 1972 An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities* **3**, 1–8.

Baum, L. E. & Petrie, T. 1966 Statistical inference from probabilistic functions of finite state Markov chain. *Ann. Math. Statist.* **37**, 1554–1563.

Baum, L. E., Petrie, T., Soules, G. & Weiss, N. 1970 A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* **41**, 164–171.

Besag, J. 1986 On the statistical analysis of dirty pictures. *Jl R. statist. Soc.* B **48**, 296–302.

Billingsley, P. 1961 *Statistical inference for Markov processes.* University of Chicago Press.

Chung, S. H., Moore, J. B., Xia, L., Premkumar, L. S. & Gage, P. W. 1990 Characterization of single channel currents using digital signal processing techniques based on Hidden Markov models. *Phil. Trans. R. Soc. Lond.* B **329**, 265–285.

Colquhoun, D. & Hawkes, A. G. 1977 Relaxation and fluctuations of membrane currents that flow through drug-operated ion channels. *Proc. R. Soc. Lond.* B **199**, 231–262.

Colquhoun, D. & Hawkes, A. G. 1981 On the stochastic properties of single ion channels. *Proc. R. Soc. Lond.* B **211**, 205–235.

Colquhoun, D. & Hawkes, A. G. 1982 On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond.* B **300**, 1–59.

Cull-Candy, S. G. & Usowicz, M. M. 1989 On the multiple-conductance single channels activated by excitatory amino acids in large cerebellar neurones of the rat. *J. Physiol., Lond.* **415**, 555–582.

Dempster, A. P., Laird, N. M. & Rubin, D. B. 1977 Maximum likelihood estimation from incomplete data via the EM algorithm. *Jl R. statist. Soc.* B **39**, 1–38.

Geman, S. & Geman, D. 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis & Machine Intelligence* **6**, 721–724.

Goutsias, J. & Mendel, J. M. 1988 Optimal simultaneous detection and estimation of filtered discrete semi-Markov chains. *IEEE Trans. Information Theory* **34** (no. 3), 551–568.

Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. 1981 Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archiv. Eur. J. Physiol.* **391**, 85–100.

Henderson, G. 1990 Complexity of 5-HT pharmacology compounded by electrophysiological data. *Trends pharmac. Sci.* **11**, 265–266.

Jahr, C. E. & Stevens, C. F. 1987 Glutamate activates multiple single channel conductances in hippocampal neurones. *Nature, Lond.* **325**, 522–525.

Juang, B. H. 1984 On the Hidden Markov Model and dynamic time warping for speech recognition – a unified view. *AT&T Tech. J.* **63** (no. 7), 1213–1243.

Kemeny, J. G. & Snell, J. L. 1960 *Finite-state Markov chains.* Princeton, New Jersey: Van Nostrand.

Krishnamurthy, V. & Moore, J. B. 1991 Signal processing of semi-Markov models with decaying states. *Proc. 30th IEEE Conference on Decision & Control, Brighton, U.K.* (In the press.)

Krishnamurthy, V., Moore, J. B. & Chung, S. H. 1991*a* On Hidden Fractal Model signal processing. *Signal Processing* **24**, 177–192.

Krishnamurthy, V., Moore, J. B. & Chung, S. H. 1991*b* Hidden Markov Model signal processing in presence of unknown deterministic interferences. *Proc. 30th IEEE Conference on Decision & Control, Brighton, U.K.* (In the press.)

Krouse, M. E., Schneider, G. T. & Gage, P. W. 1986 A large anion-selective channel has seven conductance levels. *Nature, Lond.* **319**, 58–60.

Larson, M. J. & Schubert, B. O. 1979 *Probabilistic models in*

engineering sciences, vols. 1 & 2. New York: John Wiley & Sons.

Levinson, S. E., Rabiner, L. R. & Sondhi, M. M. 1983 An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition. *The Bell System Theoret. J.* **62** (4), 1035–1074.

Liebovitch, L. S. 1989 Testing fractal and Markov models of ion channel kinetics. *Biophys. J.* **53**, 373–377.

Liebovitch, L. S. & Sullivan, J. M. 1987 Fractal analysis of a voltage-dependent potassium channel from cultured mouse hippocampal neurons. *Biophys. J.* **52**, 979–988.

McManus, O. B., Spivak, C. E., Blatz, A. L., Weiss, D. S. & Magleby, K. L. 1989 Fractal models, Markov models, and channel kinetics. *Biophys. J.* **55**, 383–385.

Paiss, O. 1990 Elimination of exponential interference from finite-length discrete signals. *IEEE Trans. ACSSP* **30** (12), 2189–2191.

Peters, B. C., Jr. & Walker, H. F. 1978 An iterative procedure for obtaining maximum-likelihood estimates of the parameters for a mixture of normal distributions. *SIAM Jl appl. Math.* **35**, 362–378.

Premkumar, L. S., Chung, S. H. & Gage, P. W. 1990*a* GABA-induced potassium channels in cultured neurons. *Proc. R. Soc. Lond.* B **241**, 153–158.

Premkumar, L. S., Gage, P. W. & Chung, S. H. 1990*b* Coupled potassium channels induced by arachidonic acid in cultured neurons. *Proc. R. Soc. Lond.* B **242**, 17–22.

Rabiner, L. R. & Juang, B. H. 1986 An introduction to Hidden Markov Models. *IEEE ASSP Mag.* **3** (1), 4–16.

Rabiner, L. R. 1989 A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proc. Inst. elect. Electron. Engrs.* **77** (2), 257–285.

Sansom, M. S. P., Ball, F. G., Kerry, C. J., McGee, R., Ramsey, R. L. & Usherwood, P. N. R. 1989 Markov, fractal, diffusion, and related models of ion channel gating. *Biophys. J.* **56**, 1229–1243.

Titterington, D. M. 1984 Recursive parameter estimation using incomplete data. *Jl R. statist. Soc.* B **45**, 37–46.

Titterington, D. M., Smith, A. F. M. & Makov, V. E. 1985 *Statistical analysis of finite mixture distributions*. Wiley Series in Probability & Mathematical Statistics. New York: John Wiley & Sons.

Zeitouni, O. & Dembo, A. 1988 Transitions of finite-state continuous-time Markov processes. *IEEE Trans. Information Theory* **34** (4), 890–893.

Zimmerman, A. L. & Baylor, D. A. 1986 Cyclic GMP-sensitive conductance of retinal rod consists of aqueous pores. *Nature, Lond.* **321**, 70–72.

Received 8 April 1991; revised 9 August 1991; accepted 5 September 1991