Characterization of single channel currents using digital signal processing techniques based on Hidden Markov Models

S. H. CHUNG[1], JOHN B. MOORE[2], LIGE XIA[2], L. S. PREMKUMAR[3] AND P. W. GAGE[3]

[1] *Research School of Biological Sciences*, [2] *Research School of Physical Sciences and* [3] *John Curtin School of Medical Research, Australian National University, G.P.O. Box 4, Canberra, ACT 2601, Australia*

CONTENTS

Introduction		266
Part I:	Theoretical background	266
	(1) A general description of the technique	266
	(2) Assumptions of the model	269
	(3) Estimation of the signal and its statistics	270
	(4) Practical implementation	273
Part II:	Extraction of known signals from noise	274
	(1) Signal extraction	274
	(2) Learning transition probabilities	275
	(3) Histograms	275
	(4) Error probabilities	275
	(5) Learning discrete-state levels	276
	(6) Evolution of conditional probabilities	277
	(7) Sensitivity to the first-order Markov model assumption	277
	(8) Sensitivity to the assumption of white noise	278
	(9) Estimation of channel kinetics	279
Part III:	Application to single channel currents	280
	(1) Amplitude of small potassium currents activated by GABA	280
	(2) Time-domain properties of channel currents	281
	(3) Interpretation of the transition matrix	281
	(4) A test of effective signal extraction	282
Discussion		283
References		284
Appendix		285
Proof of equations (1) and (2)		285
Proof of equation (4)		285
Proof of equation (9)		285

SUMMARY

Techniques for extracting small, single channel ion currents from background noise are described and tested. It is assumed that single channel currents are generated by a first-order, finite-state, discrete-time, Markov process to which is added 'white' background noise from the recording apparatus (electrode, amplifiers, etc.). Given the observations and the statistics of the background noise, the techniques described here yield *a posteriori* estimates of the most likely signal statistics, including the Markov model state transition probabilities, duration (open- and closed-time) probabilities, histograms, signal levels, and the most likely state sequence.

Using variations of several algorithms previously developed for solving digital estimation problems, we have demonstrated that: (1) artificial, small, first-order, finite-state, Markov model signals embedded in simulated noise can be extracted with a high degree of accuracy, (2) processing can detect signals that do not conform to a first-order Markov model but the method is less accurate when the background noise is not white, and (3) the techniques can be used to extract from the baseline noise single channel currents in neuronal membranes. Some studies have been included to test the validity of assuming a first-order

Phil. Trans. R. Soc. Lond. B (1990) **329**, 265–285
Printed in Great Britain

Markov model for biological signals. This method can be used to obtain directly from digitized data, channel characteristics such as amplitude distributions, transition matrices and open- and closed-time durations.

INTRODUCTION

Measurement of single channel ion currents in small patches of cell membrane has proved to be a powerful tool for studying biologically important currents. As the molecular events underlying opening and closing of a receptor-channel complex are being explored in increasing detail, it is becoming necessary to characterize single channel currents more precisely than before. If, for example, an ion channel possesses multiple conductance levels, it is of interest to determine these levels and to ascertain whether the higher levels are multiples of a small 'elementary' level (Krouse *et al.* 1986; Hunter & Giebisch 1987). Furthermore, models proposed to explain the time-dependent behaviour of ion channels depend on knowledge of transition probabilities between levels. This information is particularly difficult to obtain when the currents are very small and obscured by unavoidable background noise. Of course, the signal:noise ratio can be improved to some extent by high-frequency filtering but this method obscures and distorts fast channel current transients.

We here apply digital signal processing techniques to extract small single channel currents from noise and to compute transition probabilities from one state to another, together with other relevant signal statistics. The method is based on the assumption that the current flow through a single channel is determined by a first-order, finite-state, Markov process on which white noise is imposed. The model we adopt, commonly called by engineers a 'Hidden Markov Model' (HMM), has been widely used for a variety of numerical estimation problems, including speech processing and estimation associated with convolutional coding. The procedures for 'removing' noise from the observed sequence of data or, more precisely, for obtaining signal statistics, involve several steps. First, initial values of transition probabilities from one state to another are assigned. These can be organized as an $N \times N$ matrix, where N represents the number of possible states of the signal. A so-called forward–backward procedure is then applied to give *a posteriori* signal statistics. Next, Baum–Welch re-estimation formulae are used to re-estimate the transition probabilities and the process is repeated until convergence takes place (normally within ten iterations). The signal statistics on the last iteration are preserved.

In the theoretical section of the paper, we briefly discuss some aspects of estimation theory for HMM. For detailed mathematical accounts of the theory, the reader is referred to Baum and his colleagues (1966, 1970, 1972) and Rabiner (1989). To validate the signal processing technique, a known first-order, finite-state, Markov signal is embedded in noise generated by a resistor/patch-clamp amplifier/filter system, and the technique used to extract the signal. From the results of these simulations, it is possible to illustrate the power of the techniques, to ascertain their limits in terms of error probabilities and to test the effects of deviations from a first-order Markov process and from white noise. In the last section, the potential of the techniques for biophysical applications is illustrated. They are used to extract small ion currents through single channels from records dominated by noise. At the same time, transition probability matrices, open- and closed-time statistics, and signal estimates are constructed directly. Finally the significance of the techniques and results of the paper, and possible areas for future application are discussed.

PART I: THEORETICAL BACKGROUND

This section introduces the theoretical basis for the signal processing scheme based on HMMs. We first give a broad, general description of the theory and the strategy used in processing channel currents. Detailed, rigorous explanations of the assumptions we have made and the principles underlying our algorithms are provided in later sections and some of the proofs of the equations used are given in the Appendix.

(1) *A general description of the technique*

(i) *An overview*

Records of single channel currents from biological membranes consist of a signal generated by ions flowing through an open channel contaminated by 'background' noise from electrodes, amplifiers, etc. In some cases the signal is dominated by the noise and its characteristics cannot be measured with any certainty. To extract channel currents from the background noise, it is convenient to assume that the underlying signal sequence can be represented as a finite-state, discrete-time, Markov process. The model of channel dynamics previously proposed by Colquhoun & Hawkes (1977, 1981) is based on a finite-state, continuous-time Markov process, where the state represents the distinct conformational state of the channel macromolecule and the transition matrix of the process is denoted by $\boldsymbol{Q}$. In this model, the states are aggregated and partitioned into two classes, namely open and shut states. The underlying Markov process is not directly observable but some of its properties can be deduced from the behaviour of single channel currents. The 'state' in this paper refers to the conductance level, and the transition matrix for the process is denoted by $\boldsymbol{A}$.

In simple terms, the core of the method is to determine the most likely transition that occurred between any two signal current levels to cause an observed transition between two successive data points. In practice, the signal is allowed to adopt any of N levels, normally evenly spaced for convenience. For example, if the largest signal is 1 pA, 50 current levels from 0 to 1.5 pA might be allowed (interval 0.03 pA). Possible transitions could be from any of the 50 levels at the first point to any of the 50 levels at the second point. Consider the first data point. We need to make an initial (arbitrary) assumption. For example, it may be assumed that there is an equal probability of the signal being at any of N levels, namely, $\frac{1}{50}$ if $N = 50$. The separation or distance between the measured

value and each one of N possible signal levels gives the probability that the observation derives from a particular signal level. By multiplying this probability by, in the example given above, $\frac{1}{50}$, we obtain the probability of the first data point being generated by each of the possible levels. Then, the probabilities of the second point being generated by each one of the possible levels can be calculated by making use of the computed probabilities of each of the possible levels for the first data point, the transition probabilities between each of these first levels and the possible second levels (obtained from a level (state) transition probability matrix), and the statistics of the noise without signal. This procedure is repeated for each data point in a series of data points in a record and then is repeated in a reverse (backward) direction. Probabilities obtained in the forwards and backwards directions are multiplied to give probabilities for each level for each data point. The level with the highest such probability is taken as the true signal underlying that data point. To perform these computations, it is necessary to specify transition probabilities from any one level to N possible levels. As these probabilities are unknown for channel currents, a reasonable guess is made to construct a transition matrix and estimate the signal current level probabilities based initially on this guess. Then, after the first forward and backward pass, the transition probabilities between levels measured from the estimated signal are substituted for the values initially used in the level transition probability matrix. After several such passes, normally 5–10, the estimated signal sequence and transition probability matrix become fairly constant and are accepted as a true description of the channel currents. The amplitude distribution of the signal gives the most probable signal levels. It will be shown that this process 'learns' the true transition matrix and produces a model of the signal very close to the original signal, even when the signal is very much smaller than the noise.

(ii) *An outline of the method*

Extracting the real signal from a limited set of imperfectly determined measurements is a problem that commonly occurs in science. The ordered set of numbers comprising the data $y_1, y_2, y_3, \ldots, y_T, Y_T$, may be viewed as a real world process that is corrupted by noise. The aim of any digital processing scheme is to eliminate this interfering noise to separate the true signal sequence $s_1, s_2, s_3, \ldots, s_T, S_T$, from the noise and, at the same time, to derive a signal model that explains the observed set of measurements. To do this, it is necessary to specify how the signal is different from the noise and to have an independent estimate of the noise. Indeed, the effectiveness of any signal processing strategy hinges critically on an accurate specification of the characteristics of the signal.

We have chosen the techniques of HMMs to analyse channel currents. The algorithms used rest on two basic assumptions which are given in detail in §2 of Part I. Channel currents are assumed to be generated by a first-order, finite-state, Markov process, whereas the noise that is introduced in the process of experimental measurements and corrupts the data is stochastic and memory-less (white) and Gaussian. In other words, it is assumed that the magnitude of channel currents, measured at a discrete-time k, is one of a finite number, N, of states (where N can be as few as 2 and as many as 120 in our algorithms), and the probability of being in any one of the N states at time $k+1$ depends solely on the state s_k at time k. Transition probabilities of a first-order Markov process can be represented by an $N \times N$ matrix, with its diagonal elements representing the probability of remaining in the same state at time $k+1$, given that the process is found in state s_k at time k. For illustration, consider a channel characterized by three conductance states (state 1 = closed, state 2 = partly open, state 3 = fully open). Its transition matrix is of the form:

$$\boldsymbol{A} = \begin{pmatrix} 0.998 & 0.001 & 0.001 \\ 0.005 & 0.980 & 0.015 \\ 0.020 & 0.000 & 0.980 \end{pmatrix}.$$

The first row of the matrix indicates that the channel has an equal probability of going into the partly open state and fully open state from the closed state. The second row reveals that, once the channel is in the partly open state, it is more likely to go to the fully open state ($p = 0.015$) than to the closed state ($p = 0.005$). Similarly, from the last row, the transition from the fully open state to the partly open state is forbidden. Conventionally, the elements of the first row are referred to as a_{11}, a_{12} and a_{13} and those of the second row as a_{21}, a_{22} and a_{23}, etc. Equation (7) in §3 (i) describes how the open-time histogram and the mean open-time or closed-time can be directly calculated from the matrix. This three-state example, with the matrix given above, will be used to illustrate the computational steps.

Suppose that a signal sequence can assume several state levels (3 in this example) and that the transition probabilities from one state level to another are known. Suppose, further, that the initial condition at time $k = 0$ is the closed state. The standard deviation of the noise is estimated from a segment of data points that contains no channel current. Three parameters, denoted in the subsequent sections as $\boldsymbol{\pi}$ (the initial condition), $\boldsymbol{A}$ ($N \times N$ transition matrix) and $\boldsymbol{b}(\cdot)$ (Gaussian error probability of the noise) constitute the signal model on which our entire processing scheme is based. Consider five measured data points, y_1, y_2, y_3, y_4, y_5 obtained at time $k = 1, k = 2, \ldots, k = 5$. The true signal sequence at these five discrete times must have been, for the three-state example given above, one of the following 243 (3^5) sequences:

clock time	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
state sequence 1	{1	1	1	1	1}
state sequence 2	{2	1	1	1	1}
⋮					
state sequence 243	{3	3	3	3	3}.

The problem is to determine which one of these 243 sequences is most probable, given the $T = 5$ measurements, the standard deviation of the noise in the absence of signal, the transition matrix $\boldsymbol{A}$ and the

initial condition $\boldsymbol{\pi}$. It is possible to assign to each possible signal sequence a numerical value $P(S_T, Y_T)$ that indicates the likelihood that a specific signal sequence produced the observed sequence. The key to this computation rests on the fact that, because of the memory-less properties of the noise, the term $P(S_T, Y_T)$ can be factored into two terms: the probability of observing a given signal sequence, $P(S_T)$, and the probability of obtaining the observation sequence given the signal sequence, $P(Y_T | S_T)$. The first term is simply the product of relevant transition probabilities. For the state sequence 1 above, it will be $P(S_T) = (a_{11})^5$, and for the state sequence 2, $P(S_T) = (a_{12})(a_{21})(a_{11})^3$, remembering that the channel was assumed to be closed (state 1) at time $k = 0$. The second term is obtained from the Gaussian error probabilities. For $k = 1$, the probability of observing y_1 given that the signal was at state 1 can be computed from the equation for the Gaussian distribution curve, and similarly for $k = 2$, 3, 4 and 5. This error probability is abbreviated as $b_j(y_k)$, the probability of observing y_k given that the signal at time k was at state j. The error probabilities calculated for $k = 1, \ldots, k = 5$ are multiplied to obtain the second term, $P(Y_T | S_T)$. The most likely signal sequence is the one for which the product of these two terms, $P(S_T)\,P(Y_T | S_T)$, is maximal. Clearly, direct tabulation of every possible permutation of state sequences is computationally unfeasible for large T and a more efficient procedure is required. A special algorithm, known as the Viterbi algorithm, reduces the number of computations from about $N^T T$ to $N^2 T$ (where N is the number of possible states and T is the number of data points) and selects the most likely state sequence given the model and the observation. This algorithm is discussed in §3 (iii).

For single channel currents, however, we have no prior knowledge about the number of states, state levels and transition probabilities. Procedures that can be used together to determine the probable number of states the signal sequence contains and the associated transition matrix are discussed in §§3 (i) and 3 (ii). These are known as the forward–backward procedure and Baum–Welch re-estimation formulae. For the forward–backward procedure, we assign three abstract variables, α (forward), β (backward) and γ, to each one of N states at each discrete time. Thus, in this example, $\alpha_1(1)$, $\alpha_1(2)$ and $\alpha_1(3)$ are the α (or forward) variables associated with state 1, state 2 and state 3 at time $k = 1$. The forward and backward variables are calculated recursively, using (1) and (2) in §3 (i). Remembering that $\boldsymbol{\pi}$ is the initial condition (the probabilities that the signal at time $k = 1$ is at state 1, state 2, state 3), values of α for the three state examples are calculated as follows:

$$\alpha_1(1) = \pi_1 b_1(y_1),$$

$$\alpha_1(2) = \pi_2 b_2(y_1),$$

$$\alpha_1(3) = \pi_3 b_3(y_1).$$

The first equation stated in words is that the forward variable associated with state 1 at time $k = 1$ is the product of the initial probability of the signal being at state 1 and the probability of observing y_1 given that the signal was at state 1. The next step is to calculate $\alpha_2(1)$, $\alpha_2(2)$ and $\alpha_2(3)$ for time $k = 2$, using all α's calculated for time $k = 1$. For the signal at time $k = 2$ to be at state 1, for example, it must have made one of the three possible transitions, namely, from state 1 to state 1, state 2 to state 1 and state 3 to state 1. Thus $\alpha_2(1)$ and all subsequent $\alpha_k(i)$ are the sums of three terms:

$$\alpha_2(1) = \alpha_1(1)\, a_{11}\, b_1(y_2) + \alpha_1(2)\, a_{21}\, b_1(y_2) + \alpha_1(3)\, a_{31}\, b_1(y_2).$$

The numerical value of $\alpha_2(1)$ thus embodies the initial condition $\boldsymbol{\pi}$, the past measurement y_1, the transition probabilities a_{ij} and the present measurement y_2. The backward variable β is calculated in the same way using the recursive formula given in (2).

The forward variable $\alpha_k(i)$ is the probability of the observation sequence, $y_1, y_2, \ldots, y_k$ and the signal is at state i at time k. Similarly, $\beta_k(i)$ is the probability of the future observation sequence $y_{k+1}, y_{k+2}, \ldots, y_T$, given the signal is at state i at time k. By multiplying $\alpha_k(i)$ and $\beta_k(i)$ for each one of N states and each time from $k = 1$ to $k = T$, and suitably normalizing the product (4), we obtain the final variable $\gamma_k(i)$. The state at which $\gamma_k(i)$ is the largest is the most likely state of the signal at time k.

Both the Viterbi algorithm and the forward–backward procedure rely on the fact that we have *a priori* knowledge of the number of states, state levels and transition matrix. Access to prior knowledge of the underlying Markov signal statistics would appear to be anomalous because, if they are already known there would be little point in processing the observed set of data to obtain this information, or little point in doing experiments to begin with. Application of the Baum–Welch re-estimation formulae detailed in §3 (ii), in conjunction with the backward–forward procedure, resolves this paradox. The initial probabilities in the matrix $\boldsymbol{A}$ (as well as $\boldsymbol{\pi}$ and $\boldsymbol{b}(\cdot)$), can simply be reasonable guesses. The formulae re-estimate a new set of transition probabilities in the revised matrix $\boldsymbol{A}'$, which are more consistent with the observed sequence of data than those initially provided. Using the revised matrix $\boldsymbol{A}'$ and the same set of the observation sequence Y_T, the matrix $\boldsymbol{A}'$ is revised again, and this process is repeated several times, usually 10–15. The re-estimation of the matrix is achieved by computing, for each state at each time k, a variable $\xi_k(i,j)$, which is the probability of making a transition from state i to state j at time k. The re-estimated transition probability a_{ij}, as given in (10), is simply [the total number of transitions made from state i to state j] divided by [the total number of transitions made from state i to all allowed states]. A powerful theorem stated in (13) shows that successively re-estimated sets of transition probabilities using the Baum–Welch formulae are closer to the true set than previous sets. With successive re-estimations, it is found in practice that the transition probability a_{ij} converges globally. This means that provided initial guesses are in the vicinity of the true

values, the re-estimated values will asymptotically approach the true ones. An initial guess for the three-state matrix given above, for example, could have been 0.8 for the three diagonal elements and 0.1 for all the off-diagonal ones. The re-estimated probabilities after 10 or 15 iterations would be expected to be very close to the true values.

The Viterbi algorithm and forward–backward procedure are systematic and efficient ways of selecting the most likely signal sequence, given the data and *a priori* transition probabilities. With reasonable guesses of the transition probabilities, this processing method is useful in estimating a time domain signal sequence for channel currents that are relatively large compared to the baseline noise, such as γ-aminobutyric acid (GABA)-induced chloride channels. For small channel currents, the use of the forward–backward procedure together with the Baum–Welch re-estimation formulae permits identification of the number of states and their levels. Once this information is gained, the signal sequence as well as the relevant signal statistics can readily be obtained by re-processing the same segment of data or the subsequent data segments obtained from the same patch, using either the Viterbi algorithm or the forward–backward procedure. The method is now described in more detail with mathematical justifications for the different procedures used.

(2) *Assumptions of the model*

A schematic diagram of the particular HMM we adopt is shown in figure 1*a*. We assume that the underlying kinetics of membrane ion channels are governed by a finite-state, discrete-time, first-order Markov process. Consequently, the state s_k of a channel at time k, being the ion channel current free of measurement noise, is one of a finite number N of states $q_1, q_2, \ldots, q_N$. Moreover, with S_k denoting the sequence of states $s_1, s_2, \ldots, s_k$, then the probability of being in any state q_i at time $k+1$, given knowledge of states up to time k, depends only on the state s_k at time k. Thus $P(s_{k+1} = q_i \,|\, S_k) = P(s_{k+1} = q_i \,|\, s_k)$. The state transition probabilities of passing from state level q_i at time k to state level q_j at time $k+1$, $a_{ij} = P(s_{k+1} = q_j \,|\, s_k = q_i)$, $(k \geqslant 2)$ form a state transition probability matrix $\boldsymbol{A} = (a_{ij})$. These probabilities are assumed to be invariant of k and have the properties

$$a_{ij} \geqslant 0 \quad \text{and} \quad \sum_{j=1}^{N} a_{ij} = 1 \quad \text{(for each } i\text{)}.$$

(Throughout the paper matrices and vectors are indicated by italic or sloping bold type.)

To make a connection with the theory of HMMS, we assume that the ion channel currents measured are contaminated by noise, or equivalently, that the Markov process s_k is hidden; that is, indirectly observed by noisy measurements y_k. We denote k discrete values of the observation sequence $y_1, y_2, \ldots, y_k$ by Y_k. The vector of probability functions $\boldsymbol{b}(\cdot) = (b_i(\cdot))$ where $b_i(y_k) = P(y_k \,|\, s_k = q_i)$ are assumed invariant of k, with an independence property:

$$P(y_k \,|\, s_k = q_i, S_{k-1}, Y_{k-1}) = P(y_k \,|\, s_k = q_i)$$

or, in words, noise is stochastic and memory-less. Also we assume that the initial state probability vector $\boldsymbol{\pi} = (\pi_i)$ is defined from $\pi_i = P(s_1 = q_i)$. The HMM associated with signal levels $q_1, q_2, \ldots, q_n$ is denoted $\lambda = (\boldsymbol{A}, \boldsymbol{b}(\cdot), \boldsymbol{\pi})$. Relaxing stationarity of the underlying transition probabilities is beyond the scope of this paper.

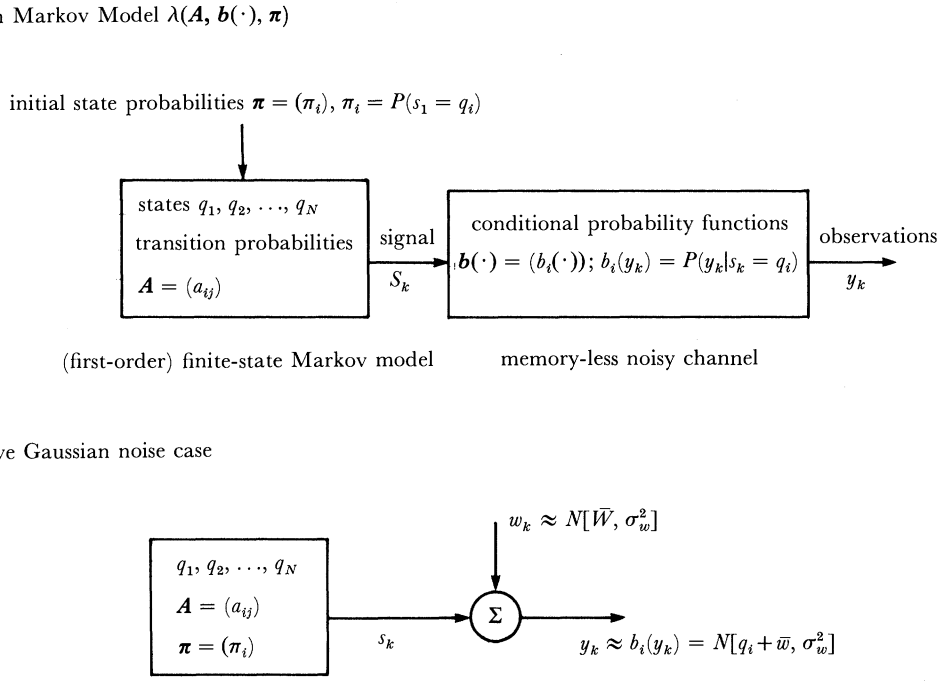


Figure 1. A block diagram of the model assumed responsible for generation of recorded single channel currents. The underlying signal sequence is generated by a discrete-time, first-order Markov process that is corrupted by memory-less noise.

In the absence of ion channel currents, experimental data give measurements y_k, which are in fact measurements of the noise generated from the pipette and amplifier. This noise is normally distributed as $N[\bar{w}, \sigma_w^2]$. Consequently, when a signal is present, it is reasonable to assume that the measurements are signal plus noise as follows:

$$y_k = s_k + w_k, \quad w_k \sim N[\bar{w}, \sigma_w^2].$$

Then

$$b_i(y_k) = (\sqrt{2\pi}\,(\sigma_w))^{-1} \exp\left\{\frac{-\frac{1}{2}|y_k - (q_i + \bar{w})|^2}{\sigma_w^2}\right\}.$$

Notice that the independence assumption of the model reflects itself here as an independence or 'whiteness' assumption of w_k. The experimental data in our case appears reasonably white over the frequency range of the bandpass (aliasing) filters (see figure 17*b*), although 'end effects' are certainly expected over the frequency range of interest. The situation is depicted in figure 1*b*. The theory and algorithms to follow do not rely on this additive normally distributed noise assumption, because re-estimation formulas can be employed to estimate $\boldsymbol{b}(\cdot)$, along with $\boldsymbol{A}$, $\boldsymbol{\pi}$, as discussed subsequently. In our computer studies we do not re-estimate $\boldsymbol{b}(\cdot)$. For simplicity, the Markov model we work with is assumed to be first-order. Extensions of the theory to mth-order Markov processes are straightforward, although the associated computations are more formidable. Aggregation (reduced state) estimation techniques can be used, as in Eyuboğlu & Qureshi (1988) and Duel-Hellen & Heegard (1989), to simplify calculations, but simplicity is gained at the expense of optimality, and also reliable statistics can only be accumulated with more observation data. Of course, the greater the observation data length, the less likely the stationarity assumption holds. Also, amplifier band-limiting filters inevitably introduce memory into the measurements, and thereby violate the first-order Markov model assumption, but we consider this memory negligible with the sampling rates and bandwidths adopted.

(3) *Estimation of the signal and its statistics*

Given the signal model as described above and given observations $y_1, y_2, \ldots, y_k$, denoted Y_k, there are four inter-related problems that can be solved.

(*a*) Evaluation of the likelihood of a given model λ generating the given sequence of data, denoted $P(Y_k|\lambda)$. This allows comparison of a set of models $\{\lambda_i\}$ to select the most likely, given the observations.

(*b*) Estimation of signal statistics such as *a posteriori* probabilities;
signal probabilities

$$\gamma_k = (\gamma_k(i)); \quad \gamma_k(i) = P(s_k = q_i | Y_T, \lambda),$$

histograms

$$\boldsymbol{h}_T = (h_T(i)); \quad h_T(i) = P(q_i | Y_T, \lambda) = \frac{1}{T}\sum_{k=1}^{T} \gamma_k(i),$$

duration-time d probabilities

$$\boldsymbol{p}_d = (p_d(i)); \quad p_d(i) = P(s_k = q_i \text{ with duration } d | Y_T, \lambda).$$

The vector γ_k represents a discrete level probability density, with $\sum_{i=1}^{N} \gamma_k(i) = 1$. This density evolves with time and displays very clearly the information which can be obtained from the data Y_T and *a priori* knowledge of the model λ. An illustrative example is studied later in figure 9. From $\gamma_k(i)$, state sequence estimates conditioned on λ, denoted $\{\hat{s}_k | \lambda\}$, such as maximum *a posteriori* (MAP) or conditional maximum (CM) estimates can be generated. Recall that MAP estimates track the peaks of the $\gamma_k(i)$ over i as k evolves, whereas the CM estimates track the means of $\gamma_k(i)$ over i as k evolves, or more precisely the nearest signal level to the mean rather than the mean itself which may not correspond to any level q_j. A measure of the quality of an estimate of time k could be taken as the variance of $\gamma_k(i)$ over i. Actually, the maximum likelihood (ML) estimate can also be independently generated by means of dynamic programming. This is the most likely path of s_k given the noisy data and a first-order Markov model λ. In computer studies presented in this paper, only MAP estimates will be used because our experience suggests that these are adequate to achieve our aims.

(*c*) Processing of the observations based on a model assumption λ and adjustment (re-estimation) of the model parameters (functions) $\boldsymbol{A}$, $\boldsymbol{b}(\cdot)$, $\boldsymbol{\pi}$, such that the updated model $\hat{\lambda}$ is more consistent with the observation sequence, and repeating until convergence to the most likely model λ^{ML} among the set $\{\lambda = (\boldsymbol{A}, \boldsymbol{b}(\cdot), \boldsymbol{\pi})\}$ given the data sequence Y_k. This then allows the estimation of the signal sequence and associated statistics for the most likely signal model in the class $\{\lambda = (\boldsymbol{A}, \boldsymbol{b}(\cdot), \boldsymbol{\pi})\}$, given the observations.

(*d*) Signal level learning. With no *a priori* knowledge of $q_1, q_2, \ldots, q_n$ then a fine quantization grid with, for example, $\sigma_w/10$ intervals can be specified and the associated model λ learnt. The associated histogram can then be used to infer quantization levels for subsequent processing.

Three HMM processing techniques reviewed more expansively in Rabiner *et al.* (1986, 1989) and depicted in figure 2 are now summarized.

(i) *Forward–backward procedure*

We consider an observation sequence Y_T of length T and an assumed signal generating model λ. In seeking knowledge about the current state s_k from the data sequence Y_T and model λ, it turns out best to first seek knowledge of s_k from past and present observations Y_k, and model λ, and in dual fashion from future observations, denoted $\bar{Y}_k$, and then combine the information appropriately.

Forward and backward vector variables $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ which summarize knowledge from past and future observations, respectively, are defined as probabilities conditioned on the model λ, for $i = 1, 2, \ldots, N$

$$\boldsymbol{\alpha}_k = (\alpha_k(i)); \quad \alpha_k(i) \triangleq P(Y_k, s_k = q_i | \lambda)$$

$$\boldsymbol{\beta}_k = (\beta_k(i)); \quad \beta_k(i) \triangleq P(\bar{Y}_k | s_k = q_i, \lambda),$$

where $\bar{Y}_k$ denotes the 'future' sequence $y_{k+1}, y_{k+2}, \ldots, y_T$, compared with Y_k which denotes the 'past' and 'present' data. We see that Y_k, $\bar{Y}_k$ together form the

observation sequence Y_T. Thus $\alpha_k(i)$ is the joint probability of the past and present observation with the present signal in state 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 formulas for $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ are readily calculated using Bayes' Rule and the first-order Markov and noise-independence assumptions (see Appendix) as

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

and

$$\beta_k(i) = \sum_{j=1}^{N} a_{ij}\, b_j(y_{k+1})\, \beta_{k+1}(j), \quad \beta_T(i) = 1. \tag{2}$$

Of course, (1) can be expressed in matrix notation as $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}'_{k-1}\boldsymbol{A}\boldsymbol{b}(y_k)$ but to continue with matrix notation would require awkward definitions and so is not attempted. Clearly the forward variable $\boldsymbol{\alpha}_k$ is calculated in a forward recursion, and the backward variable $\boldsymbol{\beta}_k$ in a backward recursion, thus the variable names and algorithm name. Because $\boldsymbol{\alpha}_k$ decays exponentially with increasing k and $\boldsymbol{\beta}_k$ with decreasing k, practical versions of these algorithms require adaptive scaling as discussed below. The evolution of $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ with k, suitably normalized, is illustrated in figure 9 for an example studied subsequently.

The likelihood function, which is the end result of the first HMM processing task, is calculated as

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

The *a posteriori* probabilities associated with the problem are given from (see Appendix)

$$\left.\begin{aligned} &\boldsymbol{\gamma}_k = (\gamma_k(i)); \\ &\gamma_k(i) = \alpha_k(i)\,\beta_k(i) \Big/ \sum_{i=1}^{N} \alpha_k(i)\,\beta_k(i), \\ &\qquad\qquad (i = 1, 2, \ldots, N). \end{aligned}\right\} \tag{4}$$

Thus $\boldsymbol{\gamma}_k$ tells us all we can know about the current state s_k, given the signal model λ and both past, present and future measurements Y_T. Its evolution is illustrated in figure 9 for an example studied subsequently.

The histogram $\boldsymbol{h}_T$ is calculated from averaging $\boldsymbol{\gamma}_k$ over all k as

$$\left.\begin{aligned} &\boldsymbol{h}_T = (h_T(i)); \\ &h_T(i) \triangleq P(q_i|Y_T, \lambda) = \frac{1}{T}\sum_{k=1}^{T} \gamma_k(i), \\ &\qquad\qquad (i = 1, 2, \ldots, N). \end{aligned}\right\} \tag{5}$$

Let us denote the sequence $\gamma_1, \gamma_2, \ldots, \gamma_k$ by Γ_k. The MAP and CM signal estimates (given λ understood) are then, respectively

$$\hat{s}_k^{\text{MAP}} = q_j \qquad \text{where } j \text{ is the value of } i \text{ which}$$

$$\text{maximizes } \gamma_k(i) \tag{6a}$$

and

$$\hat{s}_k^{\text{CM}} = q_j \qquad \text{where } j \text{ is the value of } i \text{ which}$$

$$\text{maximizes } |q_i - \sum_{h} \gamma_k(h)\, q_h|. \tag{6b}$$

Denote also $\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_k$ by $\hat{S}_k$.

The duration-time d probabilities, which are those of precisely d consecutive observations in state q_i are theoretically

$$P_i(d) = a_{ii}^{d-1}(1 - a_{ii}). \tag{7}$$

However, to get realistic estimates given Y_k and λ, the re-estimation formulas in the next section can be used. Of course the mean duration time $\bar{d}$ is then readily calculated to be as $1/(1-a_{ii})$, because

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

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

So far then, we have achieved our estimation objectives assuming a given signal model λ. We now go on to show that any estimates of λ can be improved, unless optimum (or at a local optimum). Let us define the joint conditional probabilities

$$\xi_k(i,j) = P(s_k = q_i, s_{k+1} = q_j | Y_T, \lambda). \tag{8}$$

This is readily calculated (see Appendix) as

$$\xi_k(i,j) = \alpha_k(i)\, a_{ij}\, b_j(y_{k+1})\, \beta_{k+1}(j) / P(Y_T|\lambda), \tag{9}$$

$$\hat{a}_{ij} = \sum_{k=1}^{T-1} \xi_k(i,j) \Big/ \sum_{k=1}^{T-1}\sum_{j=1}^{N} \xi_k(i,j) \tag{10}$$

and

$$\hat{\boldsymbol{\pi}}_i = \sum_{j=1}^{N} \xi_1(i,j). \tag{11}$$

In updating the vector of probability functions $\boldsymbol{b}(\cdot)$, it is reasonable to do this at a finite number of points $l_1, l_2, \ldots, l_M$ in the range of the signals y_k. Quantizing y_k to these levels, gives a quantized signal $\bar{y}_k$ and allows re-estimation of $b_i(l_j)$ as

$$\hat{b}_i(l_j) = \sum_{k=1}^{T} \gamma_k(i)\, \delta(l_j - \bar{y}_k) \Big/ \sum_{k=1}^{T} \gamma_k(i) \tag{12}$$

where $\delta(t)$ is unity when $t = 0$, and zero otherwise. A key theorem is as follows.

Theorem 1. *With the re-estimation formulas* (9)–(11) *and T sufficiently large, the model $\hat{\lambda} = \lambda(\hat{\boldsymbol{A}}, \hat{\boldsymbol{b}}(\cdot), \hat{\boldsymbol{\pi}})$ is more likely, given the observations, in that*

$$P(Y_T|\hat{\lambda}) \geqslant P(Y_T|\lambda). \tag{13}$$

The duration-time d probabilities are estimated at a discrete set of points $\{1, 2, \ldots, K\}$ as

$$\hat{P}_i(d) = \frac{\sum_{k=1}^{T} \alpha_k(i)\, P_i(d)\, \beta_{k+d}(i) \prod_{j=k+1}^{k+d} b_i(y_j)}{\sum_{d=1}^{K}\sum_{k=1}^{T} \alpha_k(i)\, P_i(d)\, \beta_{k+d}(i) \prod_{j=k+1}^{k+d} b_i(y_j)}. \tag{14}$$

The mean duration time estimate is

$$\bar{\boldsymbol{d}} = (\bar{d}_i); \quad \bar{d}_i = \sum_{d=1}^{K} P_i(d)\, d. \tag{15}$$

Proof. See for example Levinson *et al.* (1983).

Remarks.

1. The theorem is certainly not a global or even local convergence result. However, based on our simulation experience, we would expect global convergence.

2. Observe that the duration-time calculations are computationally expensive, but of course need only be performed at the last pass estimate $\hat{\lambda}$ in the schematic of figure 2.

3. When there are forbidden transitions, so that $a_{ij} = 0$ for some i,j, notice that the re-estimation formulas give $\hat{a}_{ij} = 0$ for those i,j pairs.

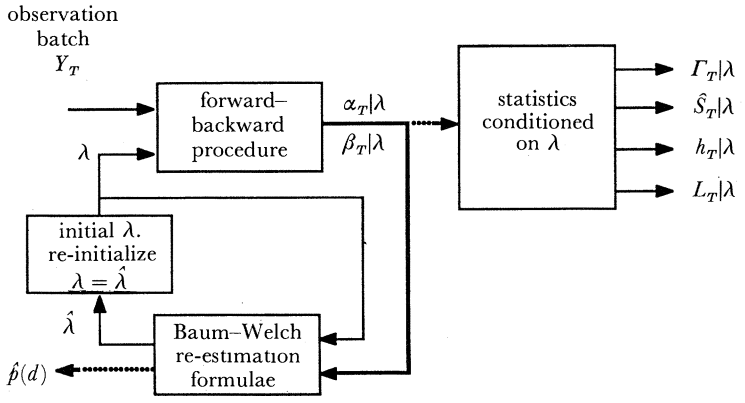
4. Empirical 'improvements' to the re-estimation formulae appear possible. For example, in updating a_{ij}, use weightings on $\xi_k(i,j)$ according to $\gamma_k(i)$, perhaps a weighting that is unity when $\gamma_k(j)$ is in the vicinity of its maximum and zero otherwise.

(iii) *Maximum likelihood signal estimation*

This subsection is not essential to appreciate the computer studies to follow, but is given here for completeness. Let us suppose that we are given the signal model λ, perhaps by means of the previous algorithms. It makes sense then to ask 'what is the most likely signal sequence?', as opposed to 'what is the best signal estimate at a particular time k?'. The second question is answered by the forward–backward procedure. The first question is answered by application of the Principle of Optimality in a Viterbi algorithm. This algorithm can efficiently yield the most likely sequence. The Principle of Optimality tells us that subsequences of optimal sequences are themselves optimal. Consequently, calculations on others need not be made. Details are now summarized.

Suppose there is given for $i = 1, 2, \ldots, N$ the most likely state estimate sequences $\hat{s}^{\mathrm{ML}}_{1\,(i)}, \hat{s}^{\mathrm{ML}}_{2\,(i)}, \ldots, \hat{s}^{\mathrm{ML}}_{k\,(i)}$, denoted $\hat{S}^{\mathrm{ML}}_{k\,(i)}$, terminating on state q_i at time k. Also, suppose there is given the associated conditional probabilities $P(\hat{S}^{\mathrm{ML}}_{k\,(i)} \mid Y_k)$ of these N state sequences given Y_k. (At $k = 1$, this information is given trivially from π.) Then, it is possible to calculate $\hat{S}^{\mathrm{ML}}_{k+1\,(i)}$ and

(*a*) multipass batch processing schematic



(*b*) on-line signal estimation

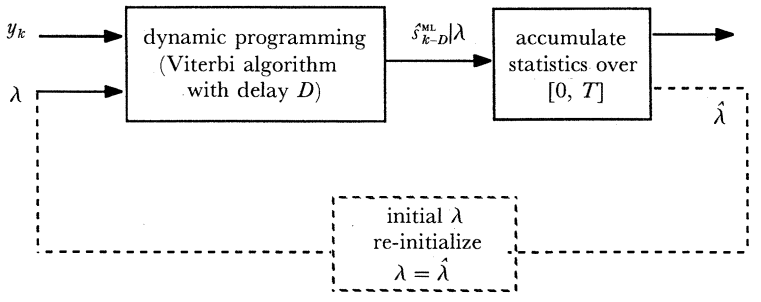


Figure 2. A block diagram showing the signal processing method. (*a*) For a segment of data, an initial transition probability matrix is arbitrarily assigned. Given the initial matrix, the algorithm based on the forward–backward procedure gives the most probable signal sequence and its statistics. With the Baum–Welch re-estimation formulae, the parameters of the model, including the transition probabilities, are adjusted, and a new signal sequence based on the adjusted parameters is estimated. This process continues until convergence takes place. (*b*) If the transition probabilities are known *a priori*, a dynamic programming method, such as the Viterbi algorithm, can be used to extract the optimal signal sequence.

$P(\hat{S}_{k+1\,(i)}^{\mathrm{ML}} \mid Y_{k+1})$ for each i as follows. For each i, consider the N transitions from state q_j at time k to state q_i at time $k+1$ for $j = 1, 2, \ldots, N$. In particular, evaluate $P(s_{k+1} = q_i, \hat{S}_{k\,(j)}^{\mathrm{ML}} \mid Y_{k+1})$ for $j = 1, 2, \ldots, N$ and select the j which maximizes this. Denote this integer m. Then set

$$S_{k+1\,(i)}^{\mathrm{ML}} = S_{k(m)}^{\mathrm{ML}}, q_m$$

$$P(\hat{S}_{k+1\,(i)}^{\mathrm{ML}} \mid Y_{(k+1)}) = P(s_{k+1} = q_m, S_{k(m)}^{\mathrm{ML}} \mid Y_{k+1}).$$

Proceeding from $k = 1$ to $k = T$ would give $\hat{S}_{T\,(i)}^{\mathrm{ML}}$ for $i = 1, 2, \ldots, N$ and $P(\hat{S}_{T\,(i)}^{\mathrm{ML}} \mid Y_T)$. Selecting the i, denoted m, which maximizes $P(\hat{S}_{T\,(i)}^{\mathrm{ML}} \mid Y_T)$ clearly gives the most likely state sequence $\hat{S}_T^{\mathrm{ML}} = \hat{S}_{T(m)}^{\mathrm{ML}}$. Actually, the paths $\hat{S}_{k\,(i)}^{\mathrm{ML}}$ tend to coalesce over the interval $[0, k-D]$ for suitably large D (say 100), so that storage requirements are simplified. In practice, it makes sense to fix some D, and store only the most likely state sequence prior to time $k-D$.

Working with logs of joint probabilities $P(S_k, Y_k) = P(S_k \mid Y_k)\, P(Y_k)$, termed lengths, rather than conditional probabilities $P(S_k \mid Y_k)$, turns out to be equivalent and simpler to implement (at least in the normally distributed noise case). Thus, using Bayes' Rule, and the first-order Markov and noise-independence assumptions, the key recursive relation is derived as

$$\begin{aligned}
&P(S_k, Y_k)\\
&\quad = P(S_k)\, P(Y_k \mid S_k)\\
&\quad = P(s_k \mid S_{k+1})\, P(S_{k-1})\, P(y_k \mid Y_{k-1}, s_k, S_{k-1})\\
&\qquad \times P(Y_{k-1} \mid s_k, S_{k-1})\\
&\quad = P(s_k \mid s_{k-1})\, P(S_{k-1})\, P(y_k \mid s_k, s_{k-1})\, P(Y_{k-1} \mid S_{k-1})\\
&\quad = [P(s_k \mid s_{k-1})\, P(y_k \mid s_k, s_{k-1})]\, P(S_{k-1}, Y_{k-1}).
\end{aligned}$$

Let us define $\xi_k(s_k, s_{k-1}) = \xi_k(i,j)$ when $s_k = q_i$ and $s_{k+1} = q_j$ and assign for each transition $\xi_k(i,j)$ from state $s_k = q_i$ at time k to state $s_{k+1} = q_j$ at time $k+1$, a 'length'

$$l[\xi_k(i,j)] \overset{\Delta}{=} -\ln(a_{ij}) - \ln P(y_{k+1} \mid \xi_k(i,j)).$$

Then the total 'length' of a sequence path on $[0, T]$

$$l_T = \sum_{k=1}^{T} l(\xi_k(s_k, s_{k-1})) = -\ln P(S_T, Y_T)$$

is minimized by the selection $\hat{S}_T^{\mathrm{ML}}$.

The Viterbi algorithm is based on the above observations and is as follows. With knowledge of $\hat{S}_{k\,(i)}^{\mathrm{ML}}$ and associated path lengths $l_k(i)$, at time $k+1$, select $\hat{S}_{k+1\,(i)}^{\mathrm{ML}}$ as the sequence $\hat{S}_{k(m)}^{\mathrm{ML}}$, $s_{k+1} = q_i$, where m is the argument such that

$$l_k(m) + l[\xi_k(m,i)] \leqslant l_k(j) + l[\xi_k(j,i)] \text{ for all } 1 \leqslant j \leqslant N.$$

That is

$$m = \underset{1 \leqslant j \leqslant N}{\operatorname{argmin}} \{l_k(j) + l[\xi_k(j,i)]\}.$$

Remarks.

1. One approach to obtaining estimates of signal statistics is to obtain first the signal estimate on-line via dynamic programming, and then take statistics of the estimated signal over a period $[0, T]$. In fact, revised transition probability estimates can be made leading to a revised estimate of λ, denoted $\hat{\lambda}$ but not the same as that of the Baum–Welch approach. The processing can be repeated with the $\hat{\lambda}$ instead of λ. The arrangement is depicted as the recursive scheme of figure 2*b*. This approach is perhaps a reasonable first cut method, but we stress that it lacks the theoretical backing that direct estimation of statistics via the forward–backward approach gives. In particular, there is no guarantee that $\hat{\lambda}$ will be an improved estimate of λ. We therefore do not explore this approach further in this paper.

2. In the case that it is known *a priori* that $a_{ij} = 0$ for certain i,j, it is clear that the most likely sequence $\hat{S}_T^{\mathrm{ML}}$ will not contain such forbidden transitions. In contrast, the estimates $\hat{S}_k^{\mathrm{MAP}}$, $\hat{S}_k^{\mathrm{CM}}$ of (6) calculated from the forward–backward procedure, when the assumed model excludes such transitions, may contain forbidden transitions. Thus it may be considered preferable to work with the forward–backward procedure together with Baum–Welch re-estimation formulae to derive signal statistics and λ^{ML}, and then use the Viterbi algorithm for obtaining a signal estimate $\hat{S}_T^{\mathrm{ML}}$, at least when it is known *a priori* that there are forbidden transitions.

(4) *Practical implementation*

In implementing the forward–backward procedure there is a 'curse of dimensionality' associated with N, and ill-conditioning due to the fact that $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ have exponential behaviour. Three specific techniques to alleviate the problems are now introduced, these being variations on techniques in the literature.

(i) *Overlapping calculations*

The 'exponential' behaviour of $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ reflects the fact that estimates at time k are influenced at an 'exponentially' decaying rate by past measurements $y_{k-1}, y_{k-2}, \ldots$, and future measurements $y_{k+1}, y_{k+2}, \ldots$, respectively. Consequently, in deriving estimates for a subinterval $[k_1, k_2]$, of $[0, T]$, then for D sufficiently large, processing observations over the interval $[k_1 - D, k_2 + D]$, and preserving estimates only in the interval $[k_1, k_2]$ will be arbitrarily close to corresponding estimates in the interval $[k_1, k_2]$ from processing the entire batch of observations in the range $[0, T]$. In our signal processing, suitable values for k_i are taken as 0, 10000, 20000, ..., points and $D = 500$ points, although reductions by a factor of 10 lead to virtually identical results.

This overlapping technique reduces both numerical ill-conditioning due to exponential behaviour, and also memory requirements, which grow linearly as the product of sub-batch length $(k_{i+1} - k_i)$ and N.

(ii) *Adaptive scaling*

The method we adopt is to monitor $\alpha_k(i)$ for each i and if the norm is less than (say) 10^{-3}, scale by a factor of 10^4. Likewise for $\beta_k(i)$. Of course, suitable records must be kept so that in calculating $\gamma(\cdot)$, $\xi(\cdot, \cdot)$, the scale factors can be included in the calculation.

(iii) *Restricted search estimation*

The computational effort of the forward–backward procedure is proportional to N^2T. It makes sense then to consider a simplified version of the algorithm which avoids calculation of very low probability transitions, given the observations. This can be achieved by considering only state transitions in the vicinity of the signal, such as within the range $[y_k-4\sigma_w, y_k+4\sigma_w]$. With say N_1 states in this range, the computational effort is of order $N_1^2\,T$.

Restrictions can be introduced in calculating $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ from (1), (2), to speed up calculations. Thus, instead of summing from over $j = 1, 2, \ldots, N$ as in (1), (2), sum over the limited range where, respectively, $\alpha_k(j)$, $\beta_{k+1}(j)$ are not negligible. It turns out that finding such a range can be achieved in a preliminary processing of the data using crude quantizations (much fewer N). Further details on this will be given in another paper. Suffice it to say that, with this approach, factors of 5 improvement in speed have been achieved in the level-learning simulation of this paper.

PART II: EXTRACTION OF KNOWN SIGNALS FROM NOISE

Before this technique was used to extract information about currents through ion channels in cell membranes, its reliability in extracting known signals buried in background noise was tested. Background noise was simulated by recording the output of the patch-clamp amplifier used in experiments with a 10 GΩ resistor representing the pipette/membrane patch combination across the input. The amplitude distributions of the noise filtered at 5 kHz, 2 kHz and 1 kHz were Gaussian. To this noise was added a first-order, finite-state, Markov signal of varied amplitude, duration and known transition probabilities. By processing the resulting observations, we have ascertained, firstly, that the technique is capable of reliably extracting signals with amplitudes as low as half the standard deviation (σ_w) of the baseline noise. As expected, the detection errors decreased as the average number of successive points $\bar{d}$ (dwell-time) at any one signal level increased, indicating a decrease in detection error for a particular pulse as pulse length increases. Secondly, for such signals, the transition probabilities estimated after each pass approached the true transition probabilities monotonically with each pass. Typically, the convergence rate appeared exponential, and the estimated and true transition probabilities were virtually identical after five to ten iterations. Thirdly, associated amplitude histograms and dwell-time probabilities could be calculated from the final transition matrix. Finally, we have deliberately violated the underlying assumptions of the processing scheme by applying it to extract a deterministic (not a first-order Markov) signal and to extract signals from band-limited (non-white) noise. The processing scheme was not very sensitive to deviation from the first-order Markov assumption, but the accuracy in extraction of signal timing decreased when the underlying noise was not white.

(1) *Signal extraction*

Noise recorded from a 10 GΩ resistor with a patch-clamp amplifier (Axopatch 1C, Axon Instruments) was filtered at 5 kHz (−3 dB, Bessel) and digitized at 11 kHz to simulate noise from a 'quiet' patch of membrane. A 1000 point segment of the noise is shown in figure 3*a*. The standard deviation of the noise was 0.16 pA. A Markov signal was generated with four possible levels at 1, 2, 3 and 4 standard deviations of the noise and transition probabilities of $a_{ii} = 0.97$ for i to i transitions and $a_{ij} = 0.03/3$ for i to j transitions. A 1000 point segment of the Markov signal is shown in figure 3*b*. This signal was added to the noise to give a record with the signal buried in noise (figure 3*c*).

The estimation algorithm assumes no knowledge of the signal sequence S_k or its transition probabilities a_{ij} but it is assumed (for simplicity in the first instance) that the signal amplitudes can adopt any of 13 levels separated by $\sigma_w/2$, half the standard deviation of the recorded noise. This is not a necessary assumption and later we will consider an heuristic approach to learning quantization (discrete-state) levels. The algorithms assume that the noise is white, with normal distribution $N[\bar{w}, \sigma_w^2]$, where $\bar{w}$, σ_w^2 are the mean and variance of the measured noise. The technique was applied to 40000 data points including the 1000 points shown in figure 3*c* with ten passes of the forward–backward procedure and application of the Baum–Welch re-estimation formulae depicted in figure 2*a*. The signal extracted from the segment of record in figure 3*c* after ten passes is shown in figure 3*d*. The estimated signal sequence $\hat{S}_k$ is virtually indistinguishable from the original signal sequence S_k, apart from a small number of very short excursions which are not picked out.

The improvement in estimation of the signal with number of passes is illustrated in figure 4. In figure 4*a* is a segment of the original signal. Differences between the estimated and recorded signals after ten, five, three and one passes are shown in figure 4*b–e* and illustrate a 'learning' process. Clearly there is only a marginal improvement in using ten passes instead of five passes.

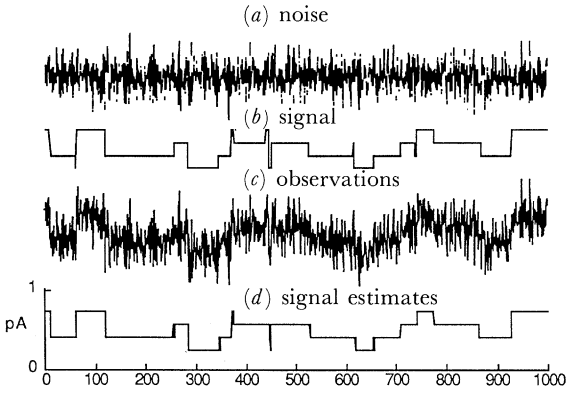


Figure 3. Extraction of a hidden signal sequence from noise. Gaussian noise (*a*) obtained from a model patch with a patch-clamp amplifier was added to a sequence of signal generated by a first-order, four-state, Markov process with $a_{ii} = 0.97$ (*b*) to give a data sequence containing the signal plus noise (*c*). The estimated signal sequence obtained by the Bayesian marginal algorithm is shown in (*d*).

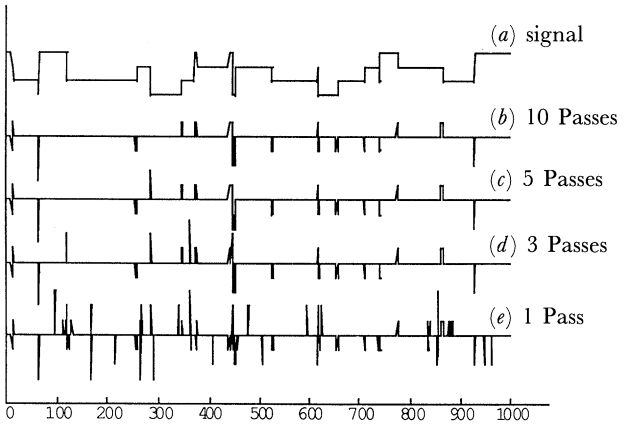


Figure 4. Estimation errors in the processing scheme. False alarms and misses were detected by subtracting the estimated signal sequence from the original signal sequence, a segment of which is shown in (*a*). (*b–e*) Differences between (*a*) and the estimated signal sequences. As the model parameters were successively adjusted, the error probabilities decreased. In the example shown here, only a marginal improvement was gained after five passes.

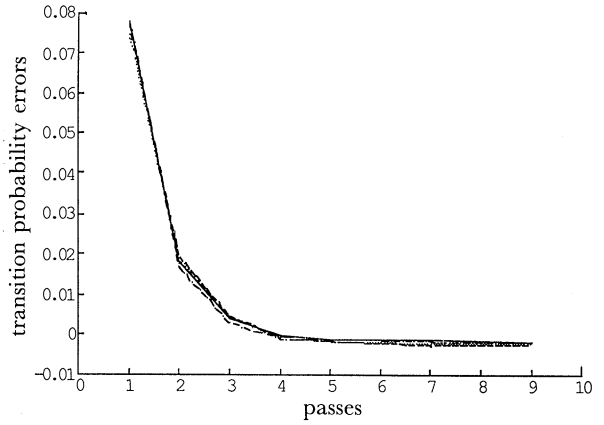


Figure 5. Estimation of the transition probabilities with the Baum–Welch re-estimation formulae. The error in the estimates of a_{ii} is plotted against the number of passes. The estimate of the transition probabilities rapidly converges to values close to, but not exactly, 0.97. The small residual discrepancies are due to the short segment of data used for re-estimation.

It can be seen that estimation errors one or two points in width occur at some transitions, and nowhere else. The signal extraction capability appears quite remarkable. It should be noted that the algorithm used 40000 data points to learn the model, rather than just the 1000 points shown in figure 3. Moreover, for simplicity in this case, the signal states are a subset of those assumed to be possible by the algorithm.

(2) *Learning transition probabilities*

As discussed in the previous section, the true transition matrix of a signal sequence we wish to extract is not known. Initially, we assign an arbitrary set of transition probabilities, and this initial matrix is revised according to the Baum–Welch formulae after each pass through a given data segment. By repeating this procedure several times, we obtain transition probabilities which are most consistent with the observation. The convergence of the estimated and true transition probabilities with number of passes and re-estimation using the Baum–Welch formulae is illustrated in figure 5. For this exercise, 20000 data points were used. The four-state signal sequence imbedded in the noise was generated from a transition matrix, $a_{ii} = 0.97$ and $a_{ij} = 0.01$ for $i \neq j$. In the initial matrix, we set $a_{ii} = 0.9$ and $a_{ij} = 0.1/3$, and the initially assigned matrix was successively revised with each pass until the values of a_{ii} converged. Note that after ten passes, there is virtually full recovery of the true transition probabilities.

(3) *Histograms*

Amplitude histograms were calculated for the 40000 data points at intervals of $\sigma_w/2$ with linear interpolation between points. The evolution of the histograms at various stages of processing of the forwards–backwards scheme with Baum–Welch re-estimation is

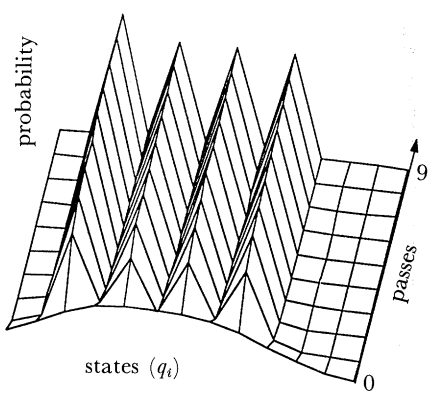


Figure 6. Gradual evolution of the signal histogram with successive re-estimation. The amplitude histogram of the original observation, indicated as 0 pass, shows no distinct peaks corresponding to the embedded signal levels but, with successive re-estimation, the peaks emerge.

depicted in figure 6. These histograms give high confidence in the extraction of signals with discrete-state levels separated by σ_w. Studies of other signal:noise ratios are given in following subsections.

It can be seen that the zero-pass histogram, being the amplitude distribution of the original data points, appears as skewed Gaussian. After one pass, there are already discernible peaks at the discrete state levels 1 σ_w, 2 σ_w, 3 σ_w and 4 σ_w, with a diminished level at the other points. There is a further concentration as the number of passes increases, with diminishing improvement after five passes. After nine passes, the probability is near zero except at 1, 2, 3 and 4 times σ_w where the probabilities are approximately equal. Thus, there appears to be a high degree of signal extraction when signal levels are spaced at intervals of σ_w.

(4) *Error probabilities*

To assess the limitations of the technique, we have made extensive comparisons of binary level, first-order Markov signals with extracted signals. In figure 7*a*, the

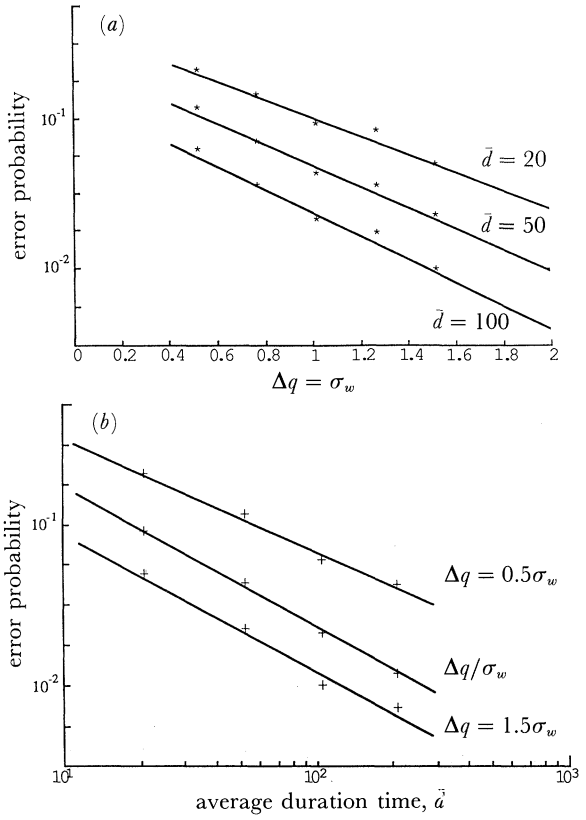


Figure 7. The probabilities of error in estimation as functions of signal amplitude (*a*) and signal duration (*b*). Markovian signals of various mean durations ($\bar{d}$; number of data points) and amplitudes (Δq) were added to noise and the signal sequences estimated. The probabilities of false alarms and misses were tabulated by subtracting the estimated signal sequence from the original signal sequence.

probability of error P_E is plotted against the separation of signal levels (quantization levels) relative to the standard deviation of the noise. Here, errors could be a 'miss' or 'false alarm'. A 'miss' is when a signal is present but not detected, and a 'false alarm' is when a signal is detected although absent. The plots are for different transition probabilities with $a_{11} = a_{22}$, $a_{12} = a_{21}$ giving different average dwell-times $\bar{d} = (1-a_{ii})^{-1}$. For low a_{ii}, there are frequent transitions so that the average signal duration $\bar{d}$ is small, and for high a_{ii}, $\bar{d}$ is relatively large. It can be seen that P_E decays 'exponentially' as the standard deviation of the noise decreases over a reasonable range. Similarly, the probability of error decays 'exponentially' with increasing $\bar{d}$ (figure 7*b*).

A more complete analysis, at least for moderately low error probabilities (say 10^{-3}), is available in Forney (1972) for the case of binary signals with $a_{11} = a_{22} = a_{21} = a_{12} = 0.5$ in normally distributed white noise. In this case, error probability bounds can be formulated in terms of the probability function

$$Q(x) = (2\pi)^{-\frac{1}{2}} \int_x^\infty \exp(-y^2/2)\,\mathrm{d}y,$$

which can be approximated for large x as

$$Q(x) \approx \frac{\exp(-x^2/2)}{[x\sqrt{2\pi}]}.$$

The error probability bounds are:

$$K_1 Q[d_{\min}/(2\sigma_w)] \leqslant P_E \leqslant K_2 Q[d_{\min}/(2\sigma_w)],$$

where $d^2_{\min}$ is the minimum energy of possible signal pulses, and K_i are given in Forney (1972). The derivation approach is not immediately applicable with $a_{ij} \neq 0.5$ as here, although we would expect behaviour to be in terms of $Q(\cdot)$ functions involving signal:noise ratios as above.

The parameters in figure 7 are expressed in terms of the ratio of signal amplitude to noise standard deviation and the mean duration is expressed as the number of data points. In practice, noise standard deviation from a typical patch in a quiescent state, when filtered at 2 kHz, is about 0.1 pA. If a digitizing interval of 100 μs is used, $d = 100$ represents 10 ms. Thus a 0.15 pA signal lasting 10 ms will be detected with a probability of error of about 1 %. The false alarm or miss is most likely to occur at the transition (cf. figure 4), so introducing on average a 100 μs error in the estimation of the duration of a 10 ms signal.

(5) *Learning discrete-state levels*

When the possible levels of a signal q_i are unknown, as is normally the case, relatively small intervals between possible levels can be assumed for the algorithm. The levels that occur can then be obtained from probability histograms such as those in figure 6. This technique is illustrated for a known, first-order, Markov signal buried in noise with standard deviation σ_w with signal levels at 0.4 σ_w, 0.6 σ_w and 0.8 σ_w and with $a_{ii} = 0.98$ for all i, and $a_{ij} = 0.02/3$ for all i to j. The algorithm assumes possible signal levels at intervals of 0.1 σ_w. The evolution of the histograms for 40000 points is illustrated in figure 8. It should be pointed out that the computational effort required is an order

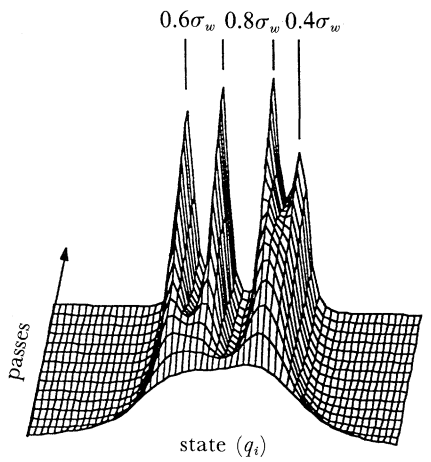


Figure 8. Learning discrete state levels. Noise was added to a four-state Markovian signal sequence, with levels at 0, 0.4, 0.6 and 0.8 of the noise standard deviation. The method was used to estimate the most likely signal levels. Amplitude histograms with successive re-estimation of the parameters are shown. The number of allowed states for this estimation was 60, at 0.1 σ_w intervals. By the fifteenth re-iteration, peaks corresponding to the original signal levels are clearly identified.

magnitude higher than for the histograms of figure 6. It can be seen that the effective noise variance decreases as the number of passes increases and after 15 passes has been decreased by an order of magnitude. Of course, the more data processed, the greater the potential for improvement of signal:noise ratio due to processing. On a finite data batch, there is clearly a resolution limit in detecting signal levels. For example, in figure 8, the resolution is greater than 0.4 σ_w for 15 passes.

Once signal levels are known, it makes sense to reprocess the data assuming only these levels to obtain a signal estimate. In this way, there will be less noise in the signal estimate. In processing more realistic data when there would be less certainty of the precise location of signal levels, it may be useful to apply a Gaussian-sum fit to the histogram. Working with reasonable variances, the means in the Gaussian-sum would indicate the discrete-state levels. One would expect that the higher the occupancy rate of a level, the less the variance of uncertainty for the Gaussian term in the sum representing this level.

(6) *Evolution of conditional probabilities*

It is of interest to display a segment of $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ and $\boldsymbol{\gamma}_k$, working at the high resolution used in the level learning example above. Figure 9 shows such segments along with the measurements y_k and MAP signal estimate $\hat{S}_k^{\text{MAP}}$, which track the 'hill-tops' of the $\boldsymbol{\gamma}_k$. Notice that the variance of $\boldsymbol{\gamma}_k$, indicated by the 'width' of the 'hills' in the figure varies with k. In particular, in the vicinity of a transition of s_k, the 'widths' are larger, as there is uncertainty as to when the transition takes place from the noisy data whereas, in the constant signal phase, there is more certainty as to the signal level. Notice also, that the 'width' of the $\boldsymbol{\alpha}_k$ hills are about the same as for the $\boldsymbol{\beta}_k$ hills, because past information is roughly equal in importance to future information. Also, the $\boldsymbol{\gamma}_k$ widths are less than those for $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$ because more information is involved.

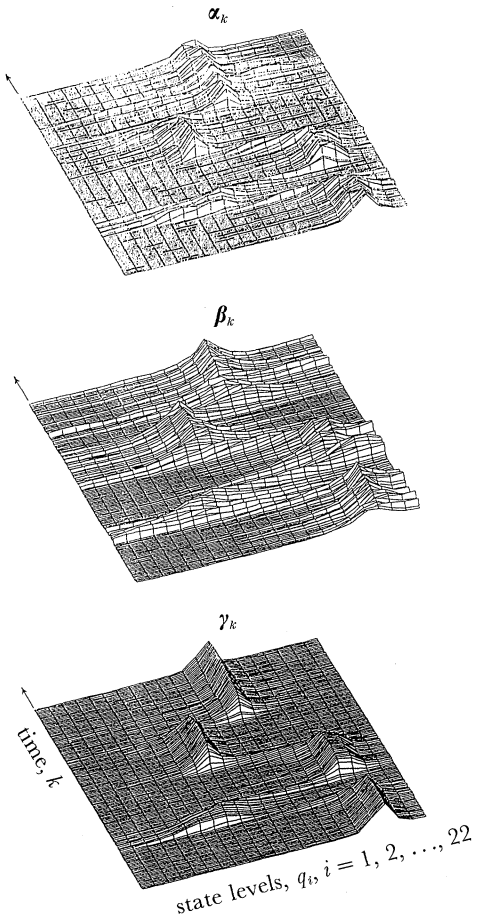


Figure 9. Plots of $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$, $\boldsymbol{\gamma}_k$ in a typical level-learning sample of 150 points taken from the level learning simulation study of figure 8. The plots were obtained during the fifteenth pass. There are $N = 22$ levels assumed for calculation giving $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$, $\boldsymbol{\gamma}_k$ vectors of order 22. The values are plotted vertically and the levels horizontally. Time is the depth ordinate. There are four levels evident in the processed data as distinct ridges. the γ plot co-ordinates the information in the α and β plots.

(7) *Sensitivity to the first-order Markov model assumption*

In the above simulations, the signal sequence embedded in the noise was first-order Markovian, generated by an $N \times N$ transition matrix. Thus the open-time distribution of the signal used in the previous sections was exponential as in (7). Typical biological signals, however, may not be best approximated by a first-order Markov process, especially if the number of events sampled is small. Here we show that the algorithm can successfully be employed in extracting signals, even when the signal statistics deviate considerably from the underlying Markov model assumption used in the estimation procedure. This suggests that the processing scheme is not very sensitive to deviation from the first-order Markov assumption, although processing may not be optimal for departures from this assumption.

Signals of four different amplitudes and a fixed duration of 9.1 ms (100 digits) were added to a segment (100000 points) of amplifier noise, filtered at 5 kHz. The standard deviation of the noise was 0.18 pA, and the amplitudes of the signals were -0.180, -0.342, -0.486 and -0.612 pA. These four levels of deterministic signal were inserted in the noise in a random order at random intervals, with the provision that signals were never superimposed. As shown in figure 10*a*, the amplitude probability density curve estimated from the signal processing scheme correctly identified all the four amplitudes of the signal. Then, using this information, we specified that the signal could only assume the five discrete states (including zero) and extracted the original signal sequence. A 1000-point segment of correctly estimated signal sequence is illustrated in figure 10*b*. The transition matrix obtained after 15 iterations describes correctly the signal statistics. The first row of the transition matrix (0.9897, 0.0029, 0.0024, 0.0029, 0.0020), gives the relative frequencies of the four different signal levels. The remaining four rows of the matrix, which read (0.01, 0.99, 0.00, 0.00, 0.00), (0.01, 0.00, 0.99, 0.00, 0.00), (0.01, 0.00, 0.00, 0.99, 0.00) and (0.01, 0.00, 0.00, 0.00, 0.99), show that the durations of

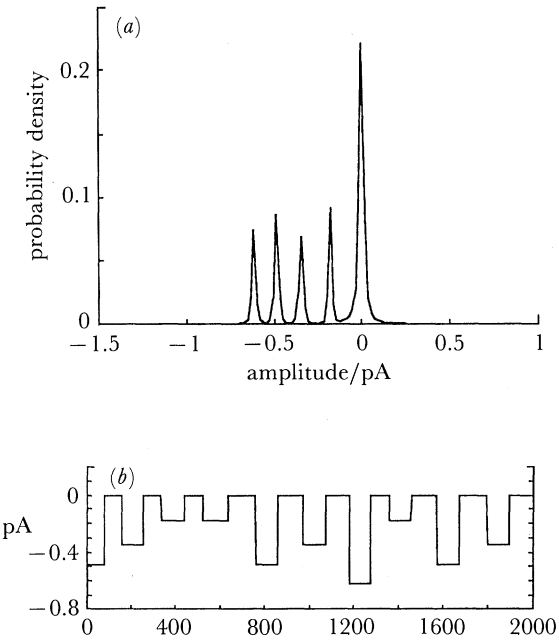


Figure 10. Extraction of deterministic signals. Signals of four different amplitudes, all 9.1 ms in duration, were added to a 100000-point segment of noise recorded from a 10 GΩ resistor with a patch-clamp amplifier. Signal levels occurred in random order at random times. The standard deviation of the noise was 0.18 pA, and the amplitudes of the signals were −0.18, −0.34, −0.48 and −0.61 pA. (*a*) The correct amplitudes of the embedded signals are identified from the probability density curve. 126 signal levels were allowed at 0.018 pA intervals. The forward–backward procedure was repeated 15 times and the transition probabilities re-estimated according to the Baum–Welch formulae. The curve shown is the final estimate of the signal amplitude probability distribution. (*b*) Using the information obtained from (*a*), the signal sequence was estimated. A 2000-point segment of estimated signal sequence shown is not distinguishable from the original fabricated signal.

all signals were 9.1 ms ($91 \times \frac{1}{0.01}$ μs) and that transitions from one open state to another did not occur.

(8) *Sensitivity to the assumption of white noise*

In the experimental situation, channel currents with added broad-band noise are filtered with a low-pass filter before sampling so that both signal and noise are distorted. To mimic the experimental procedures as closely as possible, we added periodic rectangular pulses to noise generated with a 10 GΩ resistor and patch-clamp amplifier and filtered the summed signal with a low-pass 4-pole Bessel filter. Although the underlying noise was not white within the bandwidth of the aliasing filter used in conjunction with Nyquist frequency sampling, the signal processing scheme performed reasonably well and gave reasonable estimations of the added periodic signals.

When the low-pass filter was set at 10 kHz, the noise (sampled at 22 kHz) was Gaussian with a standard deviation of 0.25 pA and its peak-to-peak fluctuation was about 1.6 pA. To this noise was added a periodic pulse of 0.1 pA with on and off durations of 5 and 10 ms, respectively and the summed noise plus signal was filtered at 2.6 or 1.3 kHz (−3 dB, 4-pole Bessel). The signal processing scheme was used as before to determine the amplitude of the pulse and the signal sequence. In all the figures illustrated in this subsection, the segments of record used for processing contained 100000 data points.

Figures 11 and 12 show segments of record filtered and sampled at different frequencies, together with amplitude probability density histograms and estimated signal sequences. In figure 11, the filter and sampling frequencies were 2.6 and 8.8 kHz; in figure 12, 1.3 and 4.4 kHz. The number of states allowed for the determination of the amplitude probability density curves were (arbitrarily) 61 (for the records filtered at 2.6 kHz) and 51 (for the records filtered at 1.3 kHz) and the interval between states was 0.01 pA.

The probability-density curves shown in figures 11*b* and 12*b* correctly show the amplitude of the added signals to be 0.10 pA. The time domain signal estimates

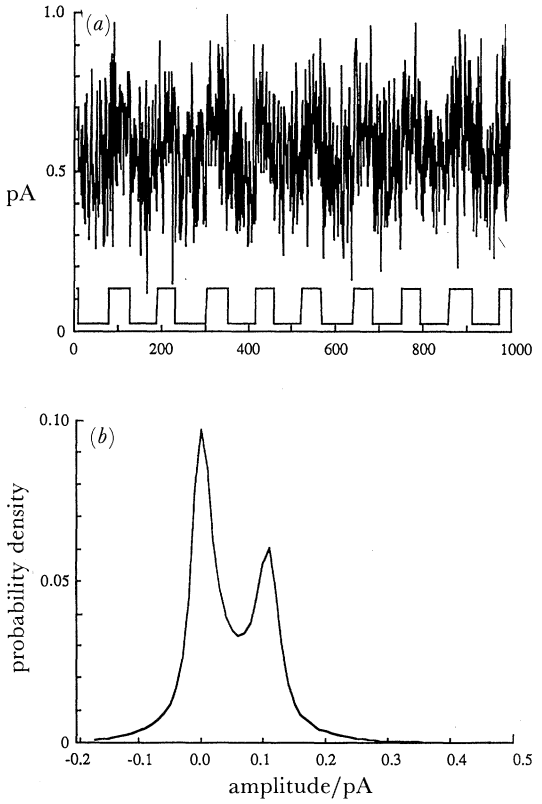


Figure 11. Extraction of small filtered periodic signals from noise. Rectangular pulses (duration 5 ms; amplitude 0.1 pA) occurring once every 15 ms were added to wide-band (cut-off frequency at 16 kHz) noise generated by a patch-clamp amplifier with a 10 GΩ resistor (model patch) across the input. The standard deviation of the noise was 0.246 pA. The record (signal plus noise) was then filtered with a 4-pole Bessel filter with a cutoff frequency of 2.6 kHz and the filtered output was digitized at 8.8 kHz (113.6 μs intervals). A 1000-point segment of the record shown in (*a*) contains the added signals. The standard deviation of noise alone filtered at 2.6 kHz was 0.14 pA. (*b*) The amplitude probability distribution was obtained by allowing 61 states separated by 0.010 pA. The separation between the peaks is 0.1 pA. The signal sequence, shown inset in (*a*), was obtained by assuming that the records contained a two-state (0 and 0.1 pA) Markovian signal.

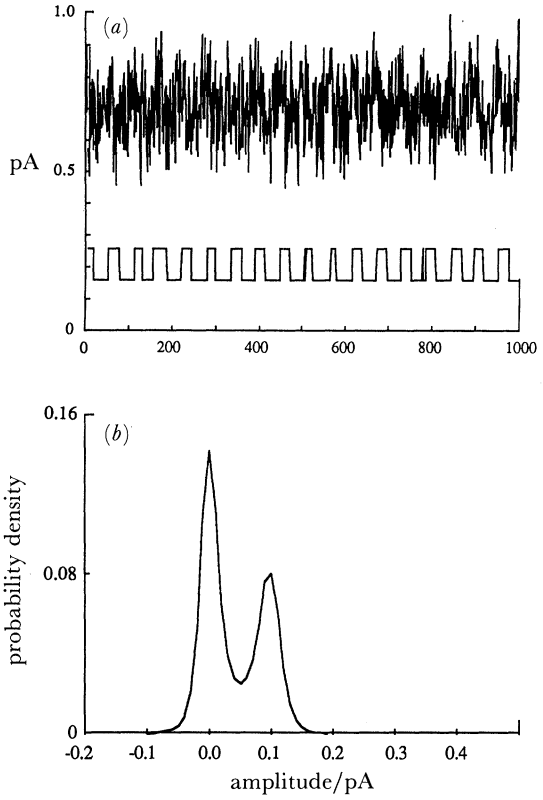


Figure 12. Identification and extraction of filtered periodic signals. The same record used in figure 11 was filtered at 1.3 kHz, and digitized at 4.4 kHz (227 μs intervals). (*a*) A 1000-point segment of the filtered summed record. (*b*) The amplitude probability density curve. A segment of estimated signal sequence is shown inset in (*a*). The standard deviation of the noise, when filtered at 1.3 kHz, was 0.09 pA. The allowed number of states for determining the amplitude histogram was 51 with a separation of 0.010 pA.

(inset in figures 11*a* and 12*a*) reveal that the extracted signal sequence mirrors the original signal more faithfully when the records are filtered at 2.6 kHz. This is in part due to the fact that the duration of the signal in terms of the number of sampling points is longer when the digitizing frequency is higher, and the error probability decreases as a function of the signal duration (cf. figure 7).

The mean duration of the extracted signal tended to be shorter than that of the original signal. We calculated the mean of the signal durations from the first 4098 points, ignoring 'false alarms' lasting one or two points. The mean duration of the estimated signals in the record filtered at 2.6 kHz and sampled at 8.8 kHz was 4.8 ± 1.1 ms; the equivalent value obtained from the record filtered at 1.3 kHz and digitized at 4.4 kHz was 4.79 ± 1.0 ms. We attribute the errors and uncertainties in the estimates of the signal durations mainly to the characteristics of the underlying noise, which was band-limited and not white. In addition, the shape of the square pulse became somewhat distorted when the signal was passed through a low-pass filter.

We have not made a systematic study of the increase in error probabilities as filter memory is gradually introduced into the noise. It is important to note, however, that the noise spectrum should optimally be flat up to the Nyquist frequency. When, for example, a record filtered at 1.3 kHz is sampled at 4.4 kHz, a short-term filter memory is introduced into the noise, and the distinction between the noise memory and Markov signal sequence becomes blurred. The signal sequence estimated under these conditions would result in a higher probability of errors (both probabilities of false alarms and misses) than if the sampling frequency were 2.6 kHz.

(9) *Estimation of single channel kinetics*

Open-time or closed-time histograms of single channel currents, from which the rate constants of the underlying channel processes can be deduced, are difficult to construct when the signal : noise ratio is low. We demonstrate here that the channel kinetics can be deduced with an acceptable degree of accuracy even

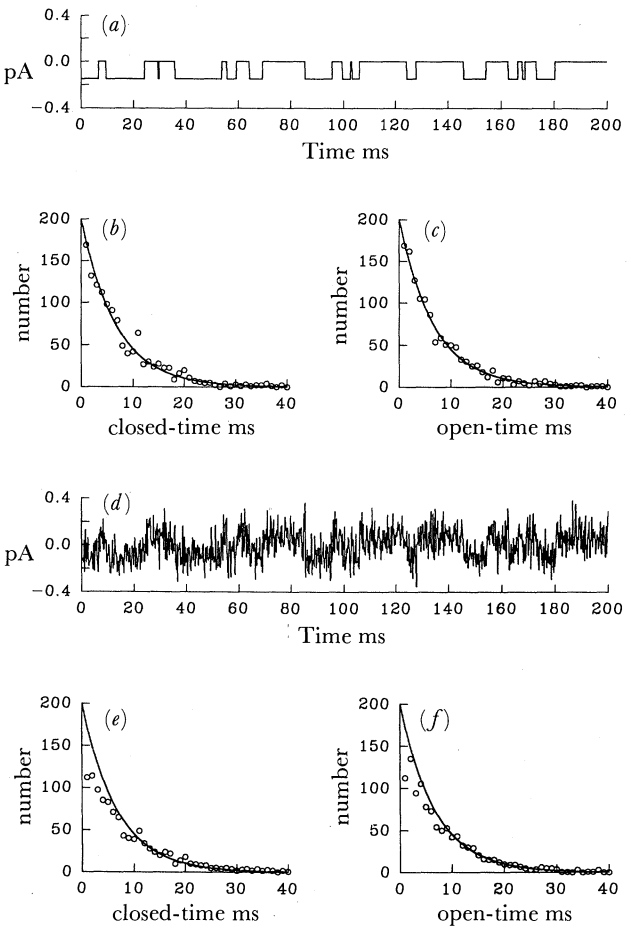


Figure 13. Determination of open- and closed-time histograms from an extracted signal sequence. The amplifier noise was filtered at 2 kHz and sampled at 5 kHz. A first-order Markovian signal with an amplitude of 0.15 pA (shown in (*a*)) was added to noise that had a standard deviation of 0.0987 pA. From the record containing the signal and noise (*d*), the signal sequence was estimated. The open-time and closed-time histograms of the original signal are shown in (*b*) and (*c*). The length of data segment used for this simulation was 100 000 points. The exponential curves fitted through the histograms of the original signal ((*b*) and (*c*)) are shown superimposed on the histograms of the extracted signal ((*e*) and (*f*)).

when current amplitudes are relatively small compared with the background noise.

A first-order, binary-state, Markovian signal sequence was added to a segment of amplifier noise. The standard deviation of the noise, when filtered at 2 kHz, was 0.0987 pA. The open-time and closed-time histograms obtained from the estimated signal sequence were then compared with those of the original sequence. The results of one such simulation, in which the amplitude of the embedded signal was 0.15 pA, are summarized in figure 13. Segments of the original signal and the signal with noise added are displayed in (*a*) and (*d*). Both open-time and closed-time histograms of the original as well as the extracted signal sequence could be fitted with single exponential curves (*b*, *c*, *e* and *f*) and are illustrated each with the same exponential curve for reference. Comparison of the two sets of histograms reveals that the mean duration deduced from the extracted signal is longer than the true value due to errors in detecting the presence of events whose durations are relatively short. The correct mean duration of the embedded signal was 7.06 ms. The estimated signal durations calculated directly from the transition matrix with the signal amplitude of 0.2 pA, 0.15 pA and 0.1 pA were, respectively, 7.17 ms (2% error), 7.4 ms (7% error) and 7.8 ms (12% error).

PART III: APPLICATION TO SINGLE CHANNEL CURRENTS

In this Part, the techniques presented above are used to extract from noise currents through ion channels in cell membranes. Currents analysed here were recorded from patches of membrane from neonatal rat pyramidal cells grown in culture for 7–10 days. The output of the patch-clamp amplifier (Axopatch 1C) was low-pass filtered at 5 kHz, digitized at 44 kHz and stored on videotape. For analysis, the digitized signal was normally sampled at 11 kHz (every 91 μs). To illustrate the application of our signal processing scheme, we have used potassium channel currents recorded from cell-attached patches on neurons exposed to GABA. These records were chosen solely to illustrate the technique; the biological significance of the observations will not be dealt with in any detail here.

As described above, noise was measured before application of the GABA and its amplitude probability curve was Gaussian. It was assumed that channel currents could be modelled as a Markov process with up to N discrete states. Initial transition matrices were arbitrarily assigned and the Baum–Welch re-estimation procedures for the parameters were repeated for 10 or 15 passes. Using the final values, the most likely signal sequence and signal statistics were derived.

(1) *Amplitude of small potassium currents activated by GABA*

Single channel potassium currents in a cell-attached patch, activated by exposing cultured hippocampal cells to 50 μM GABA, initially had a very small amplitude that was obscured by the baseline noise. The amplitude of these currents was determined by using the methods described above.

A segment (91 ms) of record obtained 90 s after the application of GABA is shown in figure 14*a*. Although there appear to be some downward signals near the middle of the trace, it is not possible to determine their amplitude by direct observation. Successive estimates of the amplitude probability density distribution of a 8000-point segment are shown in figure 14*b*. To obtain the curves, it was assumed that the signal could be at any one of 81 levels at 0.02 pA intervals. Using the techniques described above, the parameters of the signal sequence were successively adjusted after each pass. The initial amplitude distribution was a skewed Gaussian but, with further passes, three peaks emerged; the baseline at 0 pA (the largest peak), a peak at

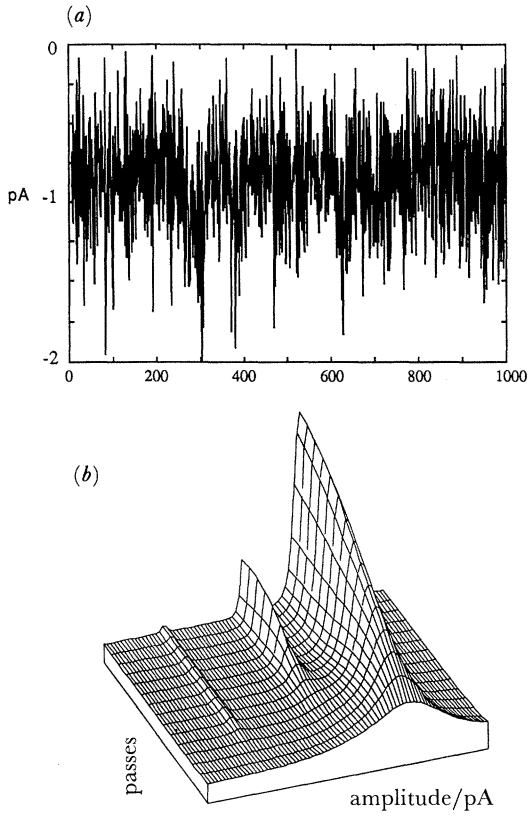


Figure 14. Determination of the amplitude of single channel potassium currents in a cell-attached patch after exposure of the cell to 50 μM GABA. (*a*) A segment of data (1000 points) recorded 90 s after the application of GABA. The record was filtered at 5 kHz and digitized at 11 kHz (91 μs intervals). In this and all subsequent figures, downward deflections represent outward currents (towards the pipette). The bath and pipette contained (in millimoles per litre): NaCl, 160; KCl, 5; $CaCl_2$, 2; $MgCl_2$, 1; HEPES, 5. The pipette potential was −40 mV with respect to the bath. (*b*) The amplitude probability density distributions were obtained from a 8000-digit segment of current record from the same patch as for (*a*). The number of allowed states was 81, ranging from +0.46 pA to −1.14 pA from the baseline. The standard deviation of the baseline noise, calculated from a segment obtained just before the application of GABA, was 0.318 pA. The distributions estimated after each of the 15 successive iterations are shown. The last distribution shows the peaks at −0.36 pA and −0.80 pA from the baseline (the highest peak).

−0.36 and another at −0.8 pA. It can be concluded that downward currents with amplitudes of about 0.36 pA and 0.8 pA were buried in the baseline noise.

(2) *Time-domain properties of channel currents*

To detemine the most likely signal sequence of the channel currents, records obtained at various times after exposure of a pyramidal cell to 100 μM GABA were analysed. The standard deviation of the noise before application of GABA was 0.25 pA. About 30 s after the application of the GABA, small downward deflections of the baseline noise could be discerned. The amplitude probability density distribution from a 17000-point segment of the record (not shown here) contained one prominent peak at −0.36 pA. Allowing signals at 0 (baseline) and −0.36 pA, a signal sequence corresponding to observed data points was then estimated. Three segments of original records and their corresponding estimated channel currents are shown in the upper row of figure 15 (*a–c*). A similar analysis of records obtained about 80 s after the agonist application showed that the amplitude of the channel currents had abruptly increased. The amplitude probability distribution obtained from a 15000-point segment of data now contained peaks at −0.72 and at −1.44 pA. Using these values as two allowed levels of the signal, we estimated sequences of channel currents. The lower row of figure 15 (*d–f*) shows three segments of the original records and estimated channel currents. The time domain signals shown are the theoretically most probable signal sequences under the assumption that the signal is a finite-state (i.e. two-state for the upper row and three-state for the lower row of figure 15), first-order Markovian which is contaminated by Gaussian noise with standard deviation of 0.25 pA.

(3) *Interpretation of the transition matrix*

The statistics of a stationary, N-state, first-order Markov signal can be characterized completely by an $N \times N$ transition matrix. Here we illustrate how transition matrices given by the techniques described above can be interpreted.

Single channel potassium currents recorded from a cell-attached patch of membrane on a cultured hippocampal neuron exposed to 100 μM GABA are used for illustration. The pipette potential was −40 mV. From a 17000-digit segment, recorded about 30 s after the application of GABA, we obtained an amplitude probability density distribution using the algorithm as before. The number of allowed states was 43 with a spacing of 0.12 pA. The amplitude histogram (not shown) revealed a single level of channel current at −0.36 pA. Using this information, we estimated a two-state signal sequence with a transition probability matrix of:

$$\boldsymbol{A} = \begin{pmatrix} 0.9984 & 0.0016 \\ 0.0448 & 0.9552 \end{pmatrix}.$$

The probabilities of the channel staying closed and going to the open state are given in the first row of the matrix, while the second row gives the probabilities of the channel closing or remaining in the open state. The mean closed and open durations are, therefore, 56.9 ms and 2.03 ms, respectively. The interval histogram would be a single exponential, as given in (7).

As the exposure to GABA was prolonged, the amplitude of the currents increased and, by 4 min after the application of GABA, single channel currents with an amplitude of about −2.5 pA were clearly seen. These currents flickered rapidly between open and partly open or closed states, and the fluctuations were

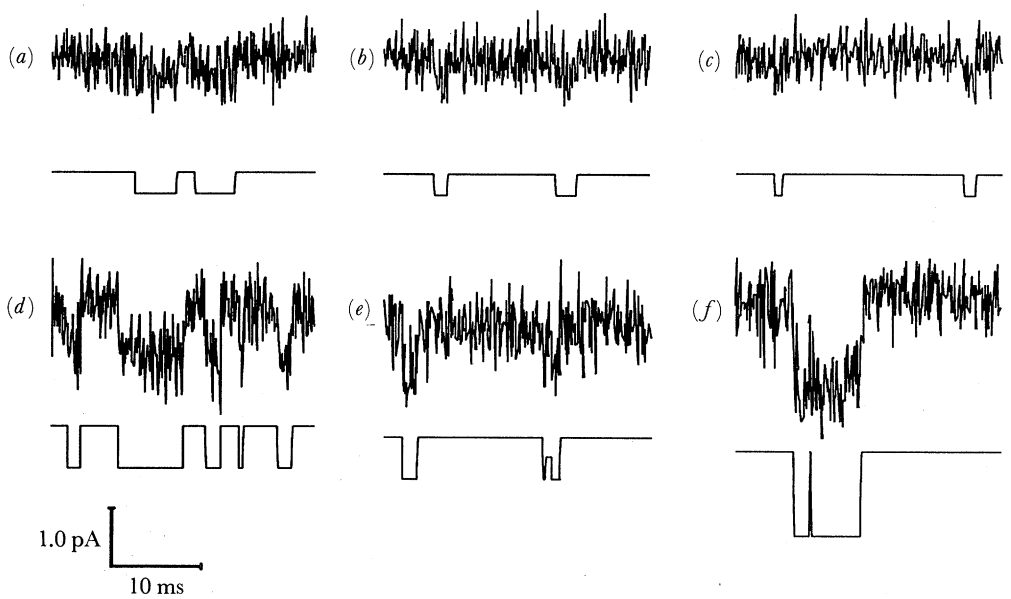


Figure 15. Records of channel activity from a cell-attached patch from a cultured hippocampal neuron. Below each trace is the most likely signal sequence. The compositions of the solutions and the pipette potential were as in figure 14. Six 300-point (27.3 ms) segments of the original records, together with the corresponding estimated signal sequences are displayed. The traces in (*a*)–(*c*) were recorded 30 s after exposure of the neuron to 100 μM GABA. Traces *d*, *e* and *f* were recorded about 1 min later. The standard deviation of the baseline noise before the application of the agonist was 0.25 pA. The amplitude probability density distributions and the estimated signal sequences were obtained from 17000-point and 15000-point records, respectively.

seen whether or not the extracellular solution contained magnesium ions. An example of these currents can be seen in figure 16*a* (note the baseline noise in this case is centred on about −1 pA). An amplitude probability density distribution and transition probability matrix (not shown) were obtained from a segment of record containing 80000 data points by allowing 43 possible

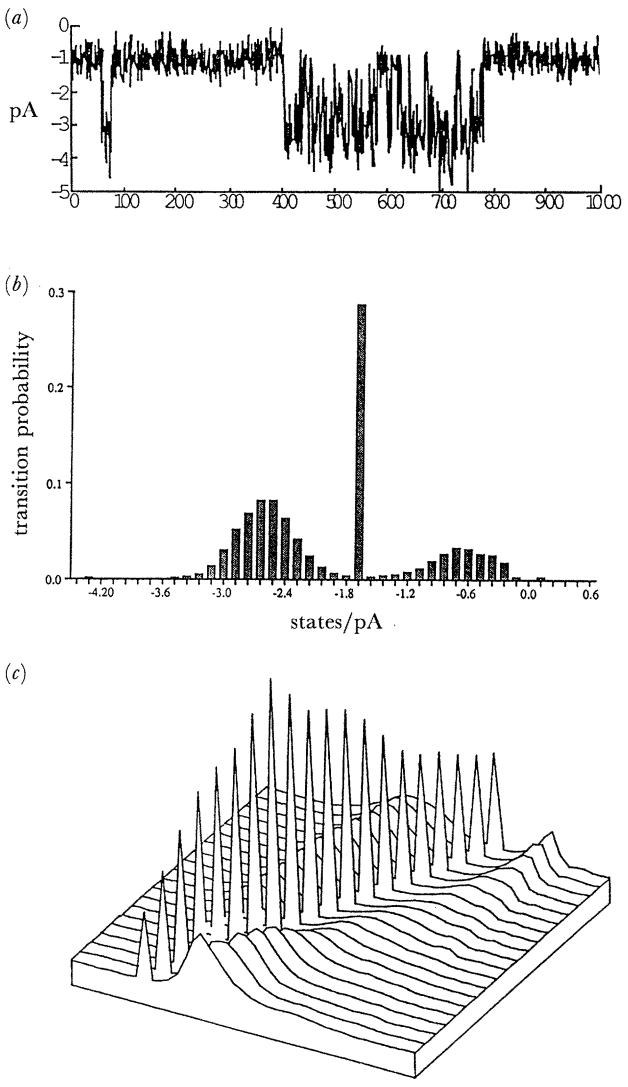


Figure 16. Currents were recorded from a cell-attached patch (pipette potential −40 mV) following exposure of the cell to 100 μM GABA. The current was filtered at 5 kHz and sampled at 11 kHz. Pipette and bath solutions were as in figure 10. (*a*) A 1000-point segment of single channel currents shows rapidly fluctuating (flickering) channel currents. (*b*) A plot of transition probability against current level (state) for one of the rows of the transition probability matrix that was obtained from a data segment containing 80000 points by allowing 43 states in steps of 0.12 pA. The highest peak in the row selected is at −1.68 pA and represents a_{ii}, the probability of remaining in the same state. The peaks on either side, one at −0.72 pA and the other at −2.52 pA, indicate that the amplitude of the channel current at −1.68 pA is most likely to increase or decrease by 0.8–0.9 pA. (*c*) Twenty of the 43 rows of the matrix are shown in a three-dimensional curve. The diagonal peaks correspond to the amplitude probability density curve. The contour of smaller peaks, running parallel to the diagonal line, indicate that changes in current amplitude are most likely to occur in steps of about 0.8–0.9 pA.

signal levels from +0.72 to −4.44 pA. As might have been expected from direct examination of the current trace, there was a single broad peak centred at −2.4 pA from the baseline peak. Examination of the transition matrix revealed that transitions from any level were generally to a level 0.8 to 0.9 pA away from that level. This is illustrated in figure 16*b* which shows transition probabilities between the −1.68 pA level and other levels. The peak at −1.68 pA is a_{ii}, the probability of remaining at −1.68 pA. The two broad peaks centred around −2.52 and −0.72 pA indicate the most probable transition levels from −1.68 pA. In figure 16*c*, 20 rows of the transition matrix are presented graphically. The first curve (lower left) shows transition probabilities from the −3.24 pA level (sharp peak); the most probable transition from this level is to the −2.4 pA level (broad peak). The curve just behind it shows transition probabilities from the −3.12 pA level; the most probable transition from this level is to the −2.28 pA level. The last curve (upper right) shows transition probabilities from the −0.84 pA level. From this state, there is an equal probability of exiting to the closed state (the 0 pA level) or to a level between −1.68 pA and −1.94 pA. From this figure, it can be concluded that the exit from an open state to the closed state is most likely to occur in small steps of about 0.8–0.9 pA. The transition from the closed state to any of the open states is also achieved in similar stepwise jumps. We should point out that these steps were obtained with the low-pass filter set at 5 kHz and sampling at 11 kHz. In another sample filtered at 10, 5 and 2 kHz, the step size was independent of filter frequency.

(4) *A test of effective signal extraction*

If channel currents contained in the original records are effectively extracted with the signal processing scheme, the residual noise obtained by subtracting the estimated signal sequence from experimentally recorded observations should be indistinguishable from control noise recorded before channels are activated. We have made several such comparisons for chloride channels activated by GABA in outside-out patches and potassium channels recorded in cell-attached patches on cells exposed to GABA or baclofen. On the basis of such tests, we conclude that the signal processing scheme adopted effectively selects all channel currents from the observed records. In all cases, the residual noise was not significantly different from the control noise.

An example of this is illustrated in figure 17. The data analysed were the same as illustrated in figure 15 (traces *d–f*). Small sections (15000 points) or the original records (top), estimated signal sequence (middle) and residual noise (bottom) are shown in figure 17*a*. Major characteristics of the residual noise and of the noise recorded just before the application of GABA were compared and found to be similar. The power spectral densities of the residual and control noise, calculated using the Maximum Entropy Method (MEM) (see Childers 1978), are shown in figure 17*b*. To obtain the spectra, each noise segment was divided into

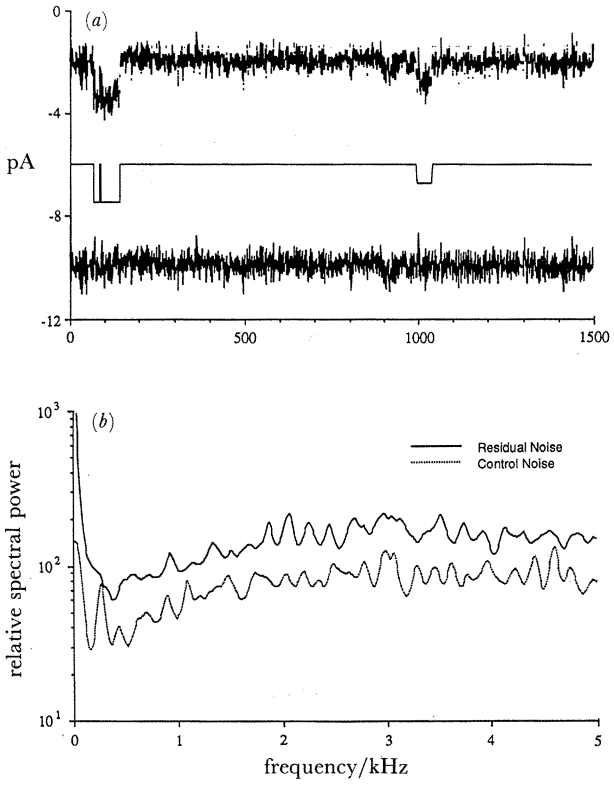


Figure 17. Comparison of control noise and residual noise. The data used for this analysis are the same as illustrated in (*d*)–(*f*) of figure 14. A 1500-point record of channel currents activated by 100 μM GABA (top), the estimated signal sequence (middle) and residual noise (bottom) are shown in (*a*). The residual noise was obtained by subtraction of the estimated signal sequence from the unprocessed current. (*b*) Power spectral densities of control noise (dotted line) and residual noise (solid line). A control segment of noise, from which the control power spectrum was calculated, was taken just before the application of GABA. The power spectra were obtained by using MEM with 64 autoregressive coefficients. The curves of both control noise and residual noise represent the average of five spectra, each calculated from a 2048-point segment.

five shorter segments containing 2048 points for calculating the autocorrelation function and power spectral density. The curves illustrated in figure 17*b* represent the averages. The autocorrelation function and MEM spectrum of the control noise (dotted line, lower trace) are virtually identical to those of the residual noise (solid line, upper trace). Both noise traces are memory-less, in that the correlation values drop to zero with one shift. Secondly, both noise signals are essentially white, as shown by the flat spectra up to 5 kHz. The skew and kurtosis of the residual noise and control noise, calculated from 10240 data points, were the same (within 1 %). The standard deviation of the residual noise was marginally higher, however, than that of the control noise (0.25 pA against 0.32 pA). This increase is due, we believe, to a small decrease in the seal resistance.

It can be concluded that the characteristics of the residual noise do not differ measurably from those of the control noise and contain no significant biological signals. Conversely, at least for the segment analysed, the parameters estimated in terms of the model were sufficient to characterize accurately the GABA-induced potassium currents.

DISCUSSION

We have described and tested a signal processing scheme based on a Hidden Markov Model for obtaining information about channel currents obscured by background noise. It is assumed that the state sequence of single channel currents is a discrete-time, first-order, finite-state, Markov process and that the observed current records containing such a signal sequence are corrupted by memory-less noise. The basic theory of this model is formulated by Baum and his colleagues (1966, 1970, 1972) and has been widely used for solving digital information problems, such as speech processing (Baker 1975; Levinson *et al.* 1983; Omura 1969). We have provided here a brief and simplified exposition of the theory. There are several reviews available that contain more detail (Rabiner 1989; Levinson *et al.* 1983; Juang 1984).

The power of the techniques for extracting known finite-state, first-order, Markov signals buried in noise generated by a model 'patch' attached to the input of the patch-clamp amplifier, has been demonstrated. The noise was assumed to be white and Gaussian when appropriately filtered at half the sampling frequency, and we have ascertained that these assumptions are reasonable for both resistor and background membrane noise (figure 17). It has been possible to recover the original sequence of the known signals with acceptable accuracy even when the signal amplitudes were as small as half the standard deviation of the background noise (figure 10). The estimates of signal model parameters and signal statistics converged 'exponentially' to the true values with successive re-estimations with the Baum–Welch formulae (figure 5). The performance of the algorithms was found to be satisfactory for detecting both Markov and non-Markov signals, although accuracy was greater for the former. Finally, even when the noise was not strictly white, the method was effective at detecting signal levels although it provided an estimate of a signal sequence that contained a higher level of errors than when the noise was white (figures 11 and 12).

When applied to records of currents generated by GABA, the technique revealed very small channel currents that could not normally be resolved (figures 14 and 15). Furthermore, the level transition matrices extracted from records that contained 'flickering' channels revealed that there was a high probability that transitions from any current level would be to levels about 0.8 pA larger or smaller (figure 16). Finally, we have illustrated how the signal processing method described here can be used to test the underlying assumptions on which the processing principle is based. The estimated signal sequence was subtracted from the original data to give the residual noise (figure 17). That the spectrum of the residual noise is virtually white and memory-less is strong evidence that the finite-state Markov process is a reasonable model to adopt for the underlying signal generating system.

The smallest signal which can be extracted with the technique depends on its duration and the standard deviation of noise. Typically, noise from a patch filtered at 5 kHz by the Axopatch-1C had a standard deviation of about 0.2 pA, and a standard deviation of about 0.1 pA when filtered at 2 kHz. However, as the filter frequency is decreased to reduce the noise, the sampling frequency needs to be decreased correspondingly, thus decreasing the duration of a given pulse length in terms of the number of digitized points and increasing the percentage error in estimating dwell-times (figure 7). Therefore care has to be taken in filtering so that the sampling frequency is not dropped too low. There is a further problem in applying the technique for determining current amplitudes. Actual current levels can be determined with closer accuracy as the allowed levels have a smaller separation. On the other hand, the probability of a false alarm increases as the separation between the allowed levels decreases. The strategy we have adopted is first to use narrowly spaced current levels to determine accurately the current levels that occur, then to reprocess the data using only those current levels to determine the signal sequence. For channel currents with a mean duration of about 5 ms, the smallest signal that can be extracted with an acceptable accuracy is about half the standard deviation of the noise.

The advantage of drastically increasing the signal:noise ratio must be weighed against the cost of processing. Typically, the forward–backward procedure requires of the order $2N_1^2\,T$ calculations, where N_1 is the number of states in the range $[-4\,\sigma_w, 4\,\sigma_w]$ and T is the number of data points. For $N_1 = 10$ and $T = 40000$, a seven MIPS computer (SUN4) can perform the C and FORTRAN implementations of the relevant calculations for one pass in about a minute. Although the computational cost is somewhat high, processing selected segments of patch-clamp data will undoubtedly provide information about details of channel currents which are buried in noise and have hitherto been largely inaccessible.

Although we have assumed that the kinetics of single channel currents can be approximated by a finite-state Markov process, we do not wish to propose that this model is the most appropriate to characterize channel currents. Colquhoun and colleagues (1977, 1981, 1983) model channel dynamics by a first-order Markov process for the underlying states of the channel macromolecules, but the number of states in general is found to be greater than the number of conductance levels. The observed current therefore is not a first-order Markov process. It is possible that the kinetics of channel currents are best described by mth order, rather than first-order, Markov processes, with or without a further constraint that there are forbidden transitions known *a priori*. One alternative model class, proposed recently by Liebovitch and his colleagues (Liebovitch *et al.* 1987; Liebovitch & Sullivan 1987), supposes that the macromolecules governing the conductance states of single channels have a continuum of many conformational states, rather than a few discrete states. Under this assumption, the transition probability from any particular state is dependent on the probabilities for leaving this collection of states and is a function of the time spent in that state. These assumptions give rise to 'fractal models'. Which one of these models provide the optimal description of observed data remains to be investigated. The procedures we have used could, with certain modifications, be applied with these additional assumptions, but the computations would be, of course, more formidable.

Many studies of ion channels are directed at constructing signal models that explain and characterize the observed data. For example, the number of conformational states a receptor-channel complex may assume is inferred from the number of exponential functions needed to fit the open- or closed-time interval distribution. Digital signal processing based on an HMM may prove to be a useful alternative tool for constructing plausible gating mechanisms that explain and characterize the observed single channel current fluctuations. The technique reduces unwanted noise considerably and at the same time generates signal statistics and the most likely model parameters. The signal estimate itself with reduced noise can give more insight into the actual signal and possible refinements to models of signal generation. The method is also useful for providing an automated, unbiased (free from observer error) and detailed description of the characteristics of larger channel currents clearly distinguishable from baseline noise. It should provide a welcome relief from the tedium of extracting information about single channel currents manually at the keyboard. Statistics such as distributions of open- and closed-times, multiple conductance levels and even the most probable signal sequence can be obtained directly and easily from amplitude histograms and the transition matrices.

We thank Professor David Colquhoun for his helpful comments on the typescript. Throughout the course of this study, Mrs Jennifer Edwards provided excellent technical assistance, for which we are grateful.

REFERENCES

Baker, J. K. 1975 The dragon system – an overview. *IEEE Trans. Acoust. Speech Signal Processing*, **ASSP-23** (No. 1), 24–29.

Baum, L. E. & Petrie, T. 1966 Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.* **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. Stat.* **41** (1), 164–171.

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

Childers, D. G. (ed.) 1987 *Modern spectrum analysis*. New York: IEEE Press.

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. & Sigworth, F. J. 1983 Fitting and statistical analysis of single-channel records. In *Single channel recording* (ed. B. Sakmann & E. Neher), pp. 191–236. New York: Plenum Press.

Duel-Hallen, A. & Heegard, C. 1989 Delayed decision-feedback sequence estimation. *IEEE Trans. Communs* **37**, 428–436.

Eyuboğlu, M. V. & Qureshi, S. V. H. 1988 Reduced-state sequence estimations with set partitioning and decision feedback. *IEEE Trans. Communs* **36**, 13–20.

Forney, G. D., Jr 1973 The Viterbi algorithm. *Proc. Inst. elect. Electron. Engrs* **61** (3), 268–278.

Hunter, M. & Giebisch, G. 1987 Multi-barrelled K channels in renal tubules. *Nature, Lond.* **322**, 522–524.

Juang, B. H. 1984 On the Hidden Markov Model and dynamic time warping for speech recognition – a unified view. *AT&T Tech. J.* **63** (7), 1213–1243.

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

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.

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

Liebovich, L. S., Fischbarg, J. & Koniaret, J. P. 1987 Ion channel kinetics: a model based on fractal scaling rather than multistate Markov process. *Mathl Biosci.* **84**, 37–68.

Omura, J. K. 1969 On the Viterbi decoding algorithm. *IEEE Trans. Inform. Theory* **IT-15**, 177–179.

Rabiner, L. R. 1986 An introduction to Hidden Markov Models. *IEEE ASSP Mag.*, 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.

(*Received 30 March 1990*; *revised 11 June 1990*; *accepted 9 July 1990*)

APPENDIX

Proof of equations (1) *and* (2)

$$\begin{aligned}\alpha_k(j) &= P(s_k = q_j, Y_k|\lambda)\\ &= \sum_{i=1}^{N} P(s_k = q_j, s_{k-1} = q_i, Y_k|\lambda)\\ &= \sum_{i=1}^{N} P(s_{k-1} = q_i, Y_{k-1}|\lambda)\\ &\quad \times P(s_k = q_j, y_k|s_{k-1} = q_i, Y_{k-1}, \lambda)\\ &= \sum_{i=1}^{N} \alpha_{k-1}(i)\, P(y_k|s_k = q_j, s_{k-1} = q_i, Y_{k-1}, \lambda)\\ &\quad \times P(s_k = q_j|s_{k-1} = q_i, Y_{k-1}, \lambda)\\ &= \sum_{i=1}^{N} \alpha_{k-1}(i)\, b_j(y_k)\, a_{ij}, \end{aligned} \tag{A 1}$$

where

$$\begin{aligned}\alpha_1(j) &= P(Y_1, s_1 = q_j|\lambda)\\ &= P(s_1 = q_j|\lambda)\, P(y_1|s_1 = q_j, \lambda)\\ &= \pi_j b_j(y_1)\\ \beta_k(i) &= P(\bar{Y}_k|s_k = q_i, \lambda)\\ &= \sum_{j=1}^{N} P(s_{k+1} = q_j, \bar{Y}_k|s_k = q_i, \lambda)\\ &= \sum_{j=1}^{N} P(\bar{Y}_{k+1}, y_{k+1}|s_{k+1} = q_j, s_k = q_i, \lambda)\\ &\quad \times P(s_{k+1} = q_j|s_k = q_i, \lambda)\\ &= \sum_{j=1}^{N} P(\bar{Y}_{k+1}|y_{k+1}, s_{k+1} = q_j, s_k = q_i, \lambda)\\ &\quad \times P(y_{k+1}|s_{k+1} = q_j, s_k = q_i, \lambda)\, a_{ij}\\ &= \sum_{j=1}^{N} \beta_{k+1}(j)\, b_j(y_{k+1})\, a_{ij}. \end{aligned} \tag{A 2}$$

Proof of equation (4)

$$\begin{aligned}\gamma_k(i) &= P(s_k = q_i|Y_T, \lambda)\\ &= P(s_k = q_i, Y_T|\lambda)\, P^{-1}(Y_T|\lambda)\\ &= P(s_k = q_i, Y_k|\lambda)\\ &\quad \times P(\bar{Y}_k|s_k = q_i, Y_k, \lambda)\, P^{-1}(Y_T|\lambda)\\ &= \alpha_k(i)\, \beta_k(i) \left[\sum_{i=1}^{N} \alpha_k(i)\, \beta_k(i)\right]^{-1} \end{aligned}$$

where

$$\begin{aligned}\sum_{i=1}^{N} \alpha_k(i)\, \beta_k(i) &= \sum_{i=1}^{N} P(Y_k, s_k = q_i|\lambda)\, P(\bar{Y}_k|s_k = q_i, \lambda)\\ &= \sum_{i=1}^{N} P(s_k = q_i, Y_k, \bar{Y}_k|\lambda)\\ &= \sum_{i=1}^{N} P(s_k = q_i, Y_T|\lambda)\\ &= P(Y_T|\lambda). \end{aligned}$$

Proof of equation (9)

$$\begin{aligned}\xi_k(i,j)&\\ &= P(s_k = q_i, s_{k+1} = q_j|Y_T, \lambda)\\ &= P(s_k = q_i, s_{k+1} = q_j, Y_T|\lambda)/P(Y_T|\lambda)\\ &= P(s_k = q_i, Y_k|\lambda)\\ &\quad \times P(s_{k+1} = q_j, \bar{Y}_{k+1}, y_{k+1}|s_k = q_i, Y_k, \lambda)/P(Y_T|\lambda)\\ &= \alpha_k(i)\, P(s_{k+1} = q_j|s_k = q_i, Y_k, \lambda)\\ &\quad \times P(\bar{Y}_{k+1}, y_{k+1}|s_k = q_i, s_{k+1} = q_j, Y_k, \lambda)/P(Y_T|\lambda)\\ &= \alpha_k(i)\, a_{ij}\, P(y_{k+1}|s_{k+1} = q_j, s_k = q_i, Y_k, \bar{Y}_{k+1}, \lambda)\\ &\quad \times P(\bar{Y}_{k+1}|s_{k+1} = q_j, s_k = q_i, Y_k, \lambda)/P(Y_T|\lambda)\\ &= \alpha_k(i)\, a_{ij}\, b_j(y_{k+1})\, \beta_{k+1}(j)/P(Y_T|\lambda). \end{aligned}$$