Superimposed HMM Transient Detection via Target Tracking Ideas

BIAO CHEN
Cornell University

PETER WILLETT
University of Connecticut

The quickest detection of superimposed hidden Markov model (HMM) transient signals is addressed. It is assumed that a known HMM is always extant but at an unknown time a second known HMM may also be present, and overlapped with the previous. Two approaches are proposed. The first treats the superimposed HMMs as a unit with an expanded state space, thus converting the problem of detecting superimposed HMMs into detection of a change in HMM, this being readily solved using a previously proposed procedure. Such an approach, though excellent in terms of performance, is not suitable for the superposition of multiple HMMs with large state dimensions due to computational complexity. A second detection scheme (based on multiple target tracking ideas) with much lower computational needs but little loss in terms of performance, is therefore developed.

Manuscript received October 21, 1999; revised October 23, 2000 and February 2, 2001; released for publication February 2, 2001.

IEEE Log No. T-AES/37/3/08564.

Refereeing of this contribution was handled by X. R. Li.

This research was supported by ONR through NUWC Newport under Contract N66604-97-M-3139.

Authors' current addresses: B. Chen, Department of Electrical Engineering, Syracuse University, Syracuse, NY; P. Willett, Electrical and Systems Engineering Department, U-157, University of Connecticut, Storrs, CT 06269, E-mail: (willett@engr.uconn.edu).

I. INTRODUCTION

Quickest detection of hidden Markov modeled (HMM) transients has been studied in [7]. The intent is to exploit the possible dependency structure of a transient signal, a structure either clear on a sample-to-sample basis, or disguised as a form of contiguity in the time or frequency domains. As a motivating example [2] we may consider periodic (or pseudoperiodic) bursts of increased variance observations; this would be audible as a train of "clicks," as in Fig. 6. Alternately, or as an overlay, we may consider a transient event to undergo a cycle of behavior, perhaps a sharp onset and tonal decay, with in this case an appropriate observation process being autoregressive (AR) coefficients taken over short blocks of raw data. In any case, appropriate modeling and use of such dependence has been shown to have significant utility [7].

The technique used in [7] for detecting a change in a (single) HMM is to extend the Page test to the case of dependent observations by treating it as repeated sequential probability ratio test (SPRT). The idea has been used by other researchers [8, 9] as applied to the identification of a change in parameters for a Poisson point process, specifically to ascertain when the character of the message-arrival rate in a communication network has undergone a change. Application of a "quickest" detection approach to detection of a HMM transient is reasonable given that the *change* must be detected before the statistics revert to their ambient descriptions. Application is feasible due to the existence of the so-called forward variable of a HMM that allows efficient calculation of the likelihood function. The techniques of [7], however, may in some cases be impractical. Specifically, attention was restricted to the detection of a switch of the observations from one HMM to another. In many practical situations the problem is considerably more complex. In the underwater acoustic scenario, for example, it often happens that the transient of interest becomes active (lurks) beneath other ongoing transients, assumed previously detected. We might, for example, be interested in the detection of a transient due to engine noise or propeller turbulence; but there are already some ambient transients of biological origin. Back to our HMM assumption, this amounts to the detection of the occurrence of a new HMM superimposed upon some existing HMMs. This is illustrated in Fig. 1: at time n_0, a new HMM starts and its samples are superimposed with those from the existing HMM to form the actual observations. The task is to detect the second HMM in the presence of the first given the superimposed observations.

A direct approach based on the procedure developed in [7] is first presented. This essentially treats the superposition of two (or more) HMMs as a single HMM with expanded state space. While

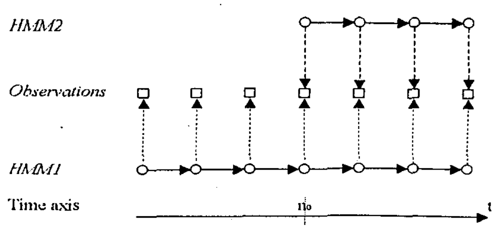


Fig. 1. Illustration of superimposed HMMs. The "□" represents either observations of HMM1 (before n_0) or superposition of observations from both HMMs (after n_0); the "o" represents underlying state of each HMM; solid arrows refer to the probabilistic evolution of the Markov states; dashed arrows refer to probabilistic realization of each HMM.

it performs very well, this direct approach can be computationally demanding due to its state space cardinality. It is reasonable, therefore, to use such a scheme when dealing with a small number of HMM transients with relatively low dimensionality. For multiple HMMs with high-dimensional state spaces, a scheme that has affordable computational complexity yet gives good detection performance is desirable.

An HMM arises in practice when there is no direct access to true underlying states, these modeled as Markov. This suggests the target tracking problem, where the state (location, velocity, etc.) is observed through noisy measurements. There is, nonetheless, significant difference between the two problems. In target tracking, the goal is to estimate. Here, our purpose is to obtain the likelihood function of the observations to be used in the detector. However, for most tracking algorithms, the output can or does include both the state estimate and the likelihood of the observation (e.g., from the innovations in linear Gaussian models). As we shall see later, inspection of the "forward variable" of a HMM reveals a close relationship with the tracking idea—the recursion of forward variables tracks the underlying state sequence of a HMM in a probabilistic sense.

For superimposed HMMs, the above naturally extends to the multiple target tracking problem. However, conventional multiple target tracking is posed as a data-association problem, and the purpose is to assign each measurement (including false alarms) to a source in either a deterministic or probabilistic sense. Such assignment is not necessarily a feature in the HMM problem. The observations are superimposed, and thus each may come from multiple HMM sources. (Superposition may be linear, as in power, or nonlinear, as by an "OR" combination of threshold exceedances.) This difference precludes direct application of existing multiple target tracking algorithms, like the Multiple Hypotheses Tracker (MHT) or Joint Probabilistic Data Association Filter (JPDAF) (see, e.g., [3, 6]). Our effort is therefore directed to the development of a new tracking algorithm that is able to infer the state sequence using the overlapped observations from multiple HMMs.

In the next section, we briefly introduce the major results of a companion paper for easy reference. Based on this we present the direct scheme for the quickest detection of superimposed HMMs using model expansion in Section III. Section IV gives the "tracking" algorithm for estimating the underlying Markov states of superimposed HMMs as well as its application to detection. Complexity analysis shows significant saving compared with the direct model expansion approach. Two examples are given in Section V to show that the tracking-based algorithm works essentially as well as the model-expansion-based Page test scheme. We deal here with the superposition of only two HMMs; extension to multiple HMMs is straightforward but notationally involved, and we do not belabor it.

II. CUSUM PROCEDURE FOR DETECTING HMMS

A. CUSUM Procedure

A standard CUSUM procedure, also known as Page's test [12], is an efficient change detection scheme. A change detection problem is such that the distribution of observations is different before and after an unknown time n_0, and we want to detect the change as soon as possible. Casting it into a standard inference framework, we have the following hypothesis testing problem

$$\begin{aligned} H: &\quad x(k) = v(k) \qquad 1 \le k \le n \\ K: &\quad x(k) = v(k) \qquad 1 \le k < n_0 \\ &\quad rm(k) = z(k) \qquad n_0 \le k \le n \end{aligned} \tag{1}$$

where $x(k)$ are observations and $v(k)$ and $z(k)$ are all independent identically distributed (IID), with probability density functions (pdf) denoted as f_H and f_K, respectively. Note that under K the observations are no longer a stationary random sequence: their distribution has a switch at n_0 from f_H to f_K.

The Page decision rule, which can be derived from the generalized likelihood ratio (GLR) test [4], amounts to finding the stopping time

$$N = \arg\min_n \left\{ \left(\max_{1\le k\le n} L_k^n \right) \ge h \right\} \tag{2}$$

where L_k^n is the log likelihood ratio (LLR) of observation x_k through x_n, and $\arg\min_n f(n)$ denotes the value of n that achieves the minimum for $f(n)$. For IID observations, (2) can be easily reformulated as [4]

$$N = \arg\min_n \left\{ \left(L(n) - \min_{1\le k\le n} L(k-1) \right) \ge h \right\} \tag{3}$$

where

$$L(k) \triangleq L_1^k = \sum_{i=1}^{k} \left(\ln \frac{f_K(x_i)}{f_H(x_i)} \right) \tag{4}$$

with $L(0) = 0$. Written in an equivalent form, a CUSUM procedure amounts to finding the following stopping time

$$N = \arg\min_n \{S_n \geq h\} \tag{5}$$

in which

$$S_n = \max\{0, S_{n-1} + g(x_n)\} \tag{6}$$

and

$$g(x_n) = \ln\left(\frac{f_K(x_n)}{f_H(x_n)}\right) \tag{7}$$

is the update nonlinearity.

Page's recursion assures that the test statistic is "clamped" at zero, i.e., whenever the LLR of current observation would make the test statistic S_n negative (which happens more often when H is true), Page's test restarts at zero. The procedure continues until it crosses the upper threshold h and a detection is claimed. There is, therefore, no false alarm rate or probability of detection involved, since we see from the implementation that, sooner or later, a detection is always claimed as long as the test is "closed" (i.e., $\Pr(N < \infty) = 1$ under both hypotheses). The performance of Page's test is therefore measured in terms of average run length (ARL) under K and H, usually denoted as D and T. Specifically, D is the avearge delay to detection and T is the ARL between two false alarms. It is always desired to have a small D while keeping T as large as possible. Analogous to the conventional hypothesis testing problem where we wish to maximize the probability of detection while keeping the false alarm rate under a fixed level, the tradeoff amounts to the choice of the upper threshold h.

As a final note, Page's test using the LLR nonlinearity has minimax optimality in terms of ARL, i.e., given a constraint on the average delay between false alarms, the Page's test minimizes the worst case delay to detection [10, 11].

B. CUSUM Procedure for Detecting HMMs

Because of the easiness in implementing a CUSUM procedure and its superior detection performance in terms of ARL, extending the procedure to dependent observations would be of practical significance; to this end, a detector reminiscent of the Page test was proposed in [1] and was shown to be optimal in Lorden's sense [10] for dependent observations. The scheme proposed in [7] is essentially the same as that in [1] except it was derived from an intuitive perspective—regarding a Page's test as repeated SPRT. Written in compact form, the Page detector for dependent observations reports upon a threshold crossing by

$$S_n = \max\{0, S_{n-1} + g(n;k)\} \tag{8}$$

where

$$g(n;k) = \ln\left(\frac{f_K(x_n \mid x_{n-1},\ldots,x_k)}{f_H(x_n \mid x_{n-1},\ldots,x_k)}\right) \tag{9}$$

and x_k is the first sample after the last reset; i.e., $S_{k-1} = 0$ and $S_j \neq 0$, for $j = k,\ldots,n-1$. Notice a direct extension would be the use of $g(n;1)$ instead of $g(n;k)$. But as pointed out in [7], the presence of dependence precludes the direct extension.

Application of the scheme to the detection of HMMs was explored by suitable use of the forward variables of an HMM, defined as

$$\alpha_t(i) = p(x_1, x_2, \ldots, x_t, s_t = i \mid \lambda) \tag{10}$$

where x_t are observations at time t, s_t is the underlying Markov state at t and λ is the HMM parameter, i.e., the state transition matrix $A = [a_{ij}]$, observation matrix $B = [b_{ij}]$ and the initial distribution of the underlying states π. It is easily checked that the following recursion holds for the forward variable

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_{jx_{t+1}}$$

with initial value

$$\alpha_1(j) = \pi(j) b_{jx_1}. \tag{11}$$

This yields an efficient way to calculate the likelihood function of an HMM given observations up to the current time. This is relevant to our approach, the on-line detection of a change. Specifically, we can write

$$f(x_1, x_2, \ldots, x_t \mid \lambda) = \sum_{i=1}^{N} \alpha_t(i). \tag{12}$$

Now the conditional probability in (9) is readily solved as

$$\begin{aligned} f_j(x_t \mid x_{t-1},\ldots,x_1) &= f(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_1, \lambda_j) \\ &= \frac{\sum_{i=1}^{N} \alpha_t(i)}{\sum_{i=1}^{N} \alpha_{t-1}(i)} \end{aligned} \tag{13}$$

where $j = H;K$. For a discussion of scaling and associated practical issues, please see [7].

III. DETECTION OF SUPERIMPOSED HMMS VIA MODEL EXPANSION

The problem of [7] is the detection of a switch from one HMM to another. For the detection of a new HMM superimposed on an existing one, as illustrated by Fig. 1, the problem is to test the following two hypotheses

$$\begin{aligned} H:&\quad y_i = z_i^1, &\quad i &= 1,2,\ldots \\ K:&\quad y_i = z_i^1, &\quad i &= 1,2,\ldots,k-1 \\ &\quad y_i = z_i^1 + z_i^2, &\quad i &= k,k+1,k+2,\ldots \end{aligned}$$

where z_i^1 is the realization of the known active HMM, denoted as HMM1, while z_i^2 is the realization of the second HMM to be detected, denoted as HMM2 (cf., Fig. 1). The "addition" in that figure is situation dependent; for example, it could be simply arithmetic addition, or it could be the result of a logical OR for binary sequences (see the examples in Section V). The unknown onset time of the second HMM is k.

Under the reasonable assumption of *independence* between the two HMMs (specifically of their hidden state sequences), the above problem is readily solved via the approach of the previous section. The idea is to consider the superposition of the two HMMs as a new HMM and apply the scheme developed for the detection of switch between two HMMs.

In reference to Fig. 1, the HMM before the change is HMM1 while afterwards it is a new HMM constructed by combining HMM1 and HMM2. The new (combined) state space is taken as the Cartesian product of the state spaces of the two superimposed HMMs, thus its dimension is the product of those of HMM1 and HMM2. Consequently, the state transition matrix A can be obtained by taking the Kronecker product of the two transition matrices, A_1 and A_2 for HMM1 and HMM2, i.e.,

$$A = A_1 \otimes A_2 = \begin{bmatrix} a_{11}^{(1)}A_2 & a_{12}^{(1)}A_2 & \cdots & a_{1n}^{(1)}A_2 \\ a_{21}^{(1)}A_2 & a_{22}^{(1)}A_2 & \cdots & a_{2n}^{(1)}A_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}^{(1)}A_2 & a_{n2}^{(1)}A_2 & \cdots & a_{nn}^{(1)}A_2 \end{bmatrix}$$

where $a_{ij}^{(1)}$ is the ith row jth column element of A_1. Construction of the observation matrix B is application specific. For example, if the observation arises from a logical OR operation on the outputs of the two constituent HMMs, then the probability of observing a "1" is the familiar sum of the probabilities of observing a 1 given the two HMMs' states, minus the product of these probabilities. The continuous-observation case for which B is a vector of pdfs can be dealt with in a similar fashion, using convolution instead of simple enumeration to construct the new observations. Note that construction of this combined HMM is off-line, and that extension to multiple HMMs is straightforward.

Once the new HMM is obtained, the technique of [7] (see Section II) for detecting a single HMM can be applied. This direct model expansion approach, as shown through numerical examples, works extremely well in terms of detection performance. Its extension to multiple superimposed HMMs, though straightforward conceptually, can be a numerical problem. To be specific, suppose we have only three superimposed HMMs with state dimensions 8, 5, 4. The resulting new HMM will have $N = 160$ possible states: each step in the forward recursion requires N^2 computations.

IV. MULTIPLE TARGET TRACKING BASED DETECTION SCHEME FOR SUPERIMPOSED HMMS

A. Motivation

The philosophy of the previous "direct" approach is to consider the individual HMMs together to form a new HMM with expanded states. Our intent here is to avoid the model-expansion step, thus obviating a possibly high-dimensional HMM; the approach builds from tracking.

Consider first the case of a single HMM. The forward variables, after suitable normalization, define the posterior probability of state occupancy given observations up to the current time. Specifically, we have

$$p(s_t = i \mid x_1, \ldots, x_t) = \frac{\alpha_t(i)}{\sum_{j=1}^{N} \alpha_t(j)} \tag{14}$$

where N is the total number of states, and we have suppressed the dependence on the HMM parameter λ. In a sense, the HMM state is being "tracked." By extension, the direct model expansion approach suggests optimal *multiple* target tracking, as we now discuss.

Let us denote $s_1(n)$ and $s_2(n)$ as the underlying states of HMM1 and HMM2 at time n, and Z_1^n as the superimposed observations $z(1)$ through $z(n)$. The goal of the tracking algorithm is obtain the likelihood function $p(Z_1^n)$ given that both HMM1 and HMM2 are active. Assume we have obtained $p(Z_1^n, s_1(n), s_2(n))$, and consider the one step update of $p(Z_1^{n+1}, s_1(n+1), s_2(n+1))$. This can be written as

$$\begin{aligned} &p(Z_1^{n+1}, s_1(n+1), s_2(n+1)) \\ &\quad = p(Z_1^n, z(n+1), s_1(n+1), s_2(n+1)) \\ &\quad = p(z(n+1) \mid s_1(n+1), s_2(n+1), Z_1^n) \\ &\qquad \times p(s_1(n+1), s_2(n+1), Z_1^n) \\ &\quad = p(z(n+1) \mid s_1(n+1), s_2(n+1)) p(s_1(n+1), s_2(n+1), Z_1^n) \end{aligned} \tag{15}$$

where the last identity follows from the fact that given $s_1(n+1)$ and $s_2(n+1)$, $z(n+1)$ is independent of the previous observations. Computation of the first term is essentially the same as obtaining the observation matrix B in the model expansion approach. For the second term in the above equation, we have

$$p(s_1(n+1), s_2(n+1), Z_1^n) = p(s_1(n+1), s_2(n+1) \mid Z_1^n) p(Z_1^n) \tag{16}$$

where $p(s_1(n+1), s_2(n+1) \mid Z_1^n)$ is usually referred to as the one step prediction probability in the tracking literature. To compute this, we use the Markov property of each individual HMM:

$$
\begin{aligned}
&p(s_1(n+1),s_2(n+1) \mid Z_1^n) \\
&\quad = \sum_{s_1(n),s_2(n)} p(s_1(n+1),s_2(n+1),s_1(n),s_2(n) \mid Z_1^n) \\
&\quad = \sum_{s_1(n),s_2(n)} p(s_1(n+1),s_2(n+1) \mid s_1(n),s_2(n),Z_1^n) \\
&\quad \times p(s_1(n),s_2(n) \mid Z_1^n) \\
&\quad = \sum_{s_1(n),s_2(n)} p(s_1(n+1) \mid s_1(n))p(s_2(n+1) \mid s_2(n)) \\
&\quad \times p(s_1(n),s_2(n) \mid Z_1^n). \qquad (17)
\end{aligned}
$$

Note that in (16) both $p(Z_1^n)$ and the posterior state probability $p(s_1(n),s_2(n) \mid Z_1^n)$ (considered as a state tracker output in a probabilistic sense) could be obtained from the previous step, i.e., from $p(Z_1^n,s_1(n),s_2(n))$, via appropriate marginalization and normalization. At any rate, having obtained the update $p(Z_1^{n+1},s_1(n+1),s_2(n+1))$, the likelihood function given both HMMs are active can then be obtained by marginalizing out $s_1(n+1)$ and $s_2(n+1)$.

This is in fact an alternative road to the model expansion method to obtain the likelihood function to be used by the detector. Its complexity is the same as that of direct model expansion as can be seen from (18), hence it is of little use from an implementational point of view. However, it serves to illustrate the relationship to tracking, and this forms a basis for the algorithm of the following section.

B. Tracking Algorithm

To reduce the computational requirements, we present a tracking algorithm built on a weak independence assumption; that is, we assume the two state sequences $\{s_1\}$ and $\{s_2\}$ are independent conditioned on previous observations, according to (18). Assumption (18) is not true; it turns out however, the output likelihood is still a good approximation to the true likelihood function under the direct independence assumption.

The assumed conditional independence of the two state sequences is

$$
p(s_1(n),s_2(n) \mid Z_1^n) = p(s_1(n) \mid Z_1^n)p(s_2(n) \mid Z_1^n) \qquad (18)
$$

and the state prediction probability is therefore

$$
\begin{aligned}
p(s_1(n+1) \mid Z_1^n) &= \sum_{s_1(n)} p(s_1(n+1) \mid s_1(n))p(s_1(n) \mid Z_1^n) \\
&\qquad (19) \\
p(s_2(n+1) \mid Z_1^n) &= \sum_{s_2(n)} p(s_1(n+1) \mid s_2(n))p(s_2(n) \mid Z_1^n)
\end{aligned}
$$

which requires $N_1^2 + N_2^2$ operations. The joint probability of $s(n+1)$ and Z_1^n is

$$
\begin{aligned}
p(s_1(n+1),Z_1^n) &= p(s_1(n+1) \mid Z_1^n)p(Z_1^n) \\
p(s_2(n+1),Z_1^n) &= p(s_2(n+1) \mid Z_1^n)p(Z_1^n)
\end{aligned} \qquad (20)
$$

which requires $N_1 + N_2$ operations. One can now calculate the joint probability of $s(n+1)$ and Z_1^{n+1}, according to

$$
\begin{aligned}
&p(s_1(n+1),Z_1^{n+1}) \\
&\quad = \sum_{s_2(n+1)} p(Z_1^n,z(n+1),s_1(n+1),s_2(n+1)) \\
&\quad = \sum_{s_2(n+1)} p(z(n+1) \mid Z_1^n,s_1(n+1),s_2(n+1)) \\
&\quad \times p(s_2(n+1) \mid s_1(n+1),Z_1^n)p(s_1(n+1),Z_1^n) \\
&\quad = \sum_{s_2(n+1)} p(z(n+1) \mid s_1(n+1),s_2(n+1)) \\
&\quad \times p(s_2(n+1) \mid Z_1^n)p(s_1(n+1),Z_1^n) \qquad (21)
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&p(s_2(n+1),Z_1^{n+1}) \\
&\quad = \sum_{s_1(n+1)} p(z(n+1) \mid s_1(n+1),s_2(n+1)) \\
&\quad \times p(s_1(n+1) \mid Z_1^n)p(s_2(n+1),Z_1^n). \qquad (22)
\end{aligned}
$$

The above requires $N_1^2N_2 + N_1N_2^2$ operations. From these we can update the posterior probability of each state

$$
\begin{aligned}
p(s_1(n+1) \mid Z_1^{n+1}) &= \frac{p(s_1(n+1),Z_1^{n+1})}{\sum_{s_1(n+1)} p(s_1(n+1),Z_1^{n+1})} \\
&\qquad (23) \\
p(s_2(n+1) \mid Z_1^{n+1}) &= \frac{p(s_2(n+1),Z_1^{n+1})}{\sum_{s_2(n+1)} p(s_2(n+1),Z_1^{n+1})}
\end{aligned}
$$

which requires $2N_1 + 2N_2$ operations. The likelihood function of the observation (which after all is the goal here) can be obtained either from (21) or (22), according to

$$
p(Z_1^{n+1}) = \sum_{s_1(n+1)} p(s_1(n+1),Z_1^{n+1}) \qquad (24)
$$

$$
p(Z_1^{n+1}) = \sum_{s_2(n+1)} p(s_2(n+1),Z_1^{n+1}) \qquad (25)
$$

requiring $\min(N_1,N_2)$ operations.

A diagram of the above tracking algorithm is shown in Fig. 2; extension to more than two HMMs is straightforward.

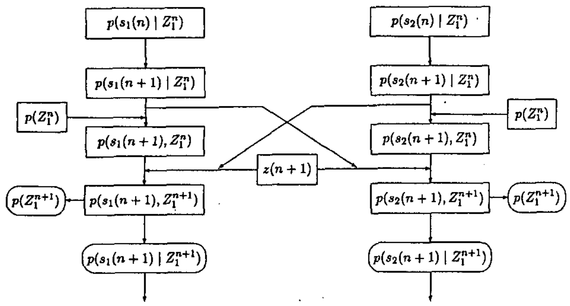


Fig. 2. Diagram of one update of tracking algorithm. Probabilities inside round boxes are outputs at each step. Likelihood $p(Z_1^{n+1})$ can be computed using either $p(s_1(n+1), Z_1^{n+1})$ or $p(s_2(n+1), Z_2^{n+1})$ as they yield essentially the same results.

C. Initialization

Initialization of the tracking algorithm involves computing $p(z(1))$ and $p(s_1(1), z(1))$, $p(s_2(1), z(1))$. Clearly, we have

$$p(z(1), s_1(1), s_2(1)) = p(z(1) | s_1(1), s_2(1)) p(s_1(1)) p(s_2(1)) \quad (26)$$

where $p(s_1(1))$ and $p(s_2(1))$ are computed using the stationary distribution obtained from each transition matrix of the two HMMs. The joint probabilities $p(s_1(1), z(1))$, $p(s_2(1), z(1))$ and the likelihood function $p(z(1))$ can be obtained via marginalization of (26). In the absence of additional information it is appropriate to initialize $p(s_1(1))$ and $p(s_2(1))$ to their stationary values; computational experience has indicated that this is not a vital concern.

D. Discussion

Multiple target tracking has been studied for decades. Though new approaches continually appear, many can be categorized as based on the JPDAF [3, 5] or on the MHT [6]. There are some cases in which an existing target tracking algorithm, such as JPDAF or MHT, can be applied almost directly in detecting superimposed HMMs. For example, certain transient signals can be modeled as slowly varying, possibly *continuous* frequency lines [15]. When modeled via an HMM, the observation (in the frequency domain) is the noisy measurement of the underlying frequency which is assumed to evolve according to a Markov model. In this case, analogy between the HMMs' state estimates and the conventional multitarget tracking problem is obvious: different frequency lines produce different (noisy) frequency domain observations, and consequently association between tracks and observations is explicit. This direct applicability is a function of the way in which the observation process is formed from the HMM; at greater generality (as we attempt) the superposition of the HMMs does not take this form, and thus the tracking algorithm presented here is neither JPDAF nor MHT. The major reason is, for tracking superimposed HMMs, there is one and only one measurement available at any instant. This measurement is the "superposition" of the realizations of two or more HMMs, and hence there is no data association in an explicit sense. Operationally, however, the tracking algorithm has a flavor similar to JPDAF.

E. Complexity

From the previous analysis, it is seen that in order to obtain the likelihood function, our tracking algorithm requires about

$$N_1^2 + N_2^2 + N_1^2 N_2 + N_1 N_2^2 + 2N_1 + 2N_2 + \min(N_1, N_2) \approx N_1^2 N_2 + N_1 N_2^2$$

operations per update in comparison with the $N^2 = (N_1 N_2)^2$ in the direct expansion approach. For three HMMs, it is straightforward to verify that we would require

$$N_1^2 + N_2^2 + N_3^2 + N_1^2 N_2 N_3 + N_1 N_2^2 N_3 + N_1 N_2 N_3^2 + 2N_1 + 2N_2 + 2N_3 + \min(N_1, N_2, N_3) \approx N_1^2 N_2 N_3 + N_1 N_2^2 N_3 + N_1 N_2 N_3^2$$

while in the direct approach the computational requirement is $(N_1 N_2 N_3)^2$.

Suppose there are k HMMs each of state dimension N; then the above tracking algorithm would require $O(kN^{k+1})$ operations per update, versus $O(N^{2k})$ in the direct approach. There is a saving of the order of N^{k-1}/k.

F. Detection

Given the output of the likelihood function of the tracker as in (24) or (25), a Page-like test is easily constructed. Under H, we use forward recursion to compute the likelihood given only HMM1 is present. Under K, the target tracker is used to compute the likelihood given both HMM1 and HMM2 are present. The output likelihood functions under both H and K are used to run a sequential test and whenever the test statistic falls below zero, it is reset to zero and the procedure restarts from the next observation. This is illustrated in Fig. 3.

Algorithmically, the detector operates as follows.

1) Set $t = 0$, $l_0 = 0$, where l_t denotes the LLR at time t.

2) Set $t = t + 1$; $t_0 = t$. Under H, initialize the forward variable $\alpha_t(\cdot \mid H)$ using (11); under K, initialize a multiple target tracker using (26). Compute the likelihood function under both hypotheses.

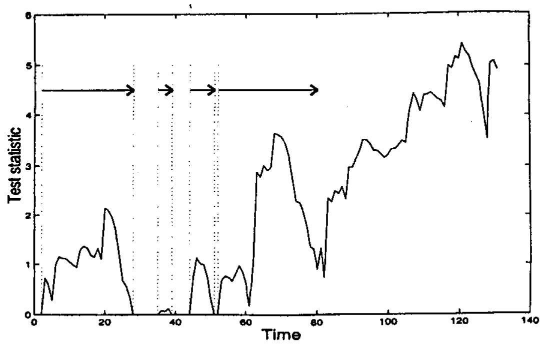


Fig. 3. An illustration of detection of proposed tracking scheme. At each tracking phase, as indicated by arrows, likelihood functions under both H and K are computed in parallel using forward variable for HMM and target tracker, respectively. If test statistic falls below zero, both forward recursion under H and target tracker under K are reinitialized and a new phase starts.

3) Update the LLR

$$l_t = l_{t-1} + \ln\left(\frac{p(Z_{t_0}^t \mid K)}{\sum_{i=1}^N \alpha_t(i \mid H)}\right). \tag{27}$$

4) If $l_t > h$, declare detection of a change, stop;
If $l_t < 0$, set $l_t = 0$; then go to 2;
If $0 < l_t < h$, continue.

5) Set $t = t + 1$; update the forward variable $\alpha_t(\cdot)$ using (13); and update the "tracker" according to (19), (20), (21), (22), and either (24) or (25); then go to 3.

If under H more than one HMM is present, then the above procedure may be modified such that the "tracking" approach is used under both hypotheses. It is often necessary to use scaled versions of the forward variables to avoid numerical underflow; please consult [7] or any standard HMM reference for details.

G. Choice of Threshold h

In Page's test, an extremely important issue is the choice of threshold. Notice this amounts to the tradeoff between the ARL under H and K. As in the conventional hypothesis testing problem where threshold is usually specified to contrain the false alarm rate, here a natural choice is to pick a threshold so that the ARL between false alarm is constrained to be greater than a certain number. To this end, it is desirable to have the relationship between T, the ARL under H, and h, the threshold. Under the IID asssumption, an asymptotic result (i.e., assuming $h \to \infty$) was obtained and it was revealed that an exponential relationship holds between T and h. For dependent observations, rigorous results in general are not available. However, we have verified in [7] that under some approximations, such an exponential relationship between T and h still holds and therefore could serve as a guideline for finding an appropriate threshold h for a prescribed T. Notice this approach applies directly only to the model expansion approach. As we see in the following section, the performance of the tracking based detection scheme is in many cases close to that of the model expansion approach. Therefore the procedure developed in [7] can also serve as an approximation to the tracking based scheme.

V. NUMERICAL EXAMPLES

The first example is a binary case, as may arise when preprocessing is simply a threshold on the signal amplitude. The second is of Gaussian increased-variance bursts: each transient is represented by a series of Gaussian bursts with increased variance and the switch between high-variance and low-variance states of the transient is assumed to follow a Markov chain (see [7] for further detail). For both cases, 50 Monte Carlo runs were conducted to obtain the ARL. For each Monte Carlo run, a sequence of 200,000 samples were generated according to the HMM models specified under both hypotheses.

A. Binary Observation HMMs

Suppose the HMM already active (extant) has four states and binary observations. The state transition and observation matrices are

$$A_1 = \begin{bmatrix} 0.990 & 0.010 & 0 & 0 \\ 0 & 0.995 & 0.005 & 0 \\ 0 & 0 & 0.995 & 0.005 \\ 0.030 & 0 & 0 & 0.970 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 0.900 & 0.100 \\ 0.900 & 0.100 \\ 0.900 & 0.100 \\ 0.300 & 0.700 \end{bmatrix}.$$

Note that in HMM1 the cascade of the first three states provides reasonably consistent 500-sample periods of mostly 0 observations, followed by a shorter burst of mostly 1 behavior. At $t = 3333$ (unknown to the detectors) a second binary-observation HMM with

$$A_2 = \begin{bmatrix} 0.995 & 0.005 \\ 0.100 & 0.900 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0.900 & 0.100 \\ 0.100 & 0.900 \end{bmatrix}$$

is superimposed; HMM2 is a more-obvious on/off switch between mostly 0 and mostly 1 observation bursts. The superposition is modeled as a logical OR, as would be reasonable if resulting from a thresholding operation. We wish to detect the onset of the second HMM as quickly as possible, and consider the following three detectors.

Detector 1 A "naive" scheme would be to count the number of 1s, since before and after the change the stationary distributions of 0 and 1 are different. This scheme treats the HMMs as if resulting from different IID sources with distinct probabilities of being 0 and 1.

In fact, this is tantamount to the Page detector which *would* be optimal (in Lorden's sense) were the observations independent. Suppose we denote the probabilities of observing a 1, respectively, before and after the onset of the second HMM as p_1 and p_2. We would then use the CUSUM

$$T_{\text{naive}}(n) = \max\{0, T_{\text{naive}}(n-1) + g(x_n)\} \tag{28}$$

in which x_n is the nth (binary) observation, and

$$g(x) = \begin{cases} \ln\left(\dfrac{p_2}{p_1}\right) & x_n = 1 \\ \ln\left(\dfrac{1-p_2}{1-p_1}\right) & x_n = 1 \end{cases}$$

is the LLR update. This detector uses no dependency information from the HMM structure.

Detector 2 We use the direct model expansion scheme. The Kronecker product of the transition matrix and the corresponding observation matrix are computed to four digits of accuracy as

$$A = \begin{bmatrix} 0.9850 & 0.0050 & 0.0100 & 0.0000 & 0 & 0 & 0 & 0 \\ 0.0990 & 0.8910 & 0.0010 & 0.0090 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9900 & 0.0050 & 0.0050 & 0.0000 & 0 & 0 \\ 0 & 0 & 0.0995 & 0.8955 & 0.0005 & 0.0045 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9900 & 0.0050 & 0.0050 & 0.0000 \\ 0 & 0 & 0 & 0 & 0.0995 & 0.8955 & 0.0005 & 0.0045 \\ 0.0298 & 0.0001 & 0 & 0 & 0 & 0 & 0.9652 & 0.0048 \\ 0.0030 & 0.0270 & 0 & 0 & 0 & 0 & 0.0970 & 0.8730 \end{bmatrix},$$

$$B = \begin{bmatrix} 0.8100 & 0.1900 \\ 0.0900 & 0.9100 \\ 0.8100 & 0.1900 \\ 0.0900 & 0.9100 \\ 0.8100 & 0.1900 \\ 0.0900 & 0.9100 \\ 0.2700 & 0.7300 \\ 0.0300 & 0.9700 \end{bmatrix}$$

Using these parameters, the detector for a change in HMM is applied.

Detector 3 The third detector uses the tracking based detector depicted in Section IV.

The results are shown in Fig. 4. Fig. 4(a) shows the number of ones in each 100 data block. Fig. 4(b) shows a typical statistic of the first detector, the "naive" method that treats the sequence as if it were of independent binary observations. Figs. 4(c) and (d) are the results of using HMM expansion and the tracking-based detection, respectively. The advantage of the latter two approaches against the IID approximation scheme is clear. To better understand the superior performance of the model-expansion-based and the tracking-based approaches as opposed to the "naive" method, the ARL between false alarms and the delay to detections of the above schemes are obtained via Monte Carlo. Essentially, by varying threshold h, the average run lengths obtained under H and K are plotted out against each other. From Fig. 5, it is clear that the performances of the model expansion based and the tracking approaches are similar, while the "naive" (independence assumption) approach is markedly inferior.

B. Gaussian Bursts with Increased Variance

Increased-variance bursts of Gaussian observations have been used to model transients that experience cycles of quiet and activity. An example is shown in Fig. 6 where Fig. 6(a) shows the transient signal and Fig. 6(b) the noisy observations within a block. Note that the observations in Fig. 6(b) are independent from a sample-to-sample point of view. However, the bursty behavior indicates possible dependence that could be utilized to improve detection performance.

Suppose a hidden Markov modeled transient is initially active, and a second starts later and is overlapped (superimposed linearly). The first HMM transient has a two-state Markov chain with transition

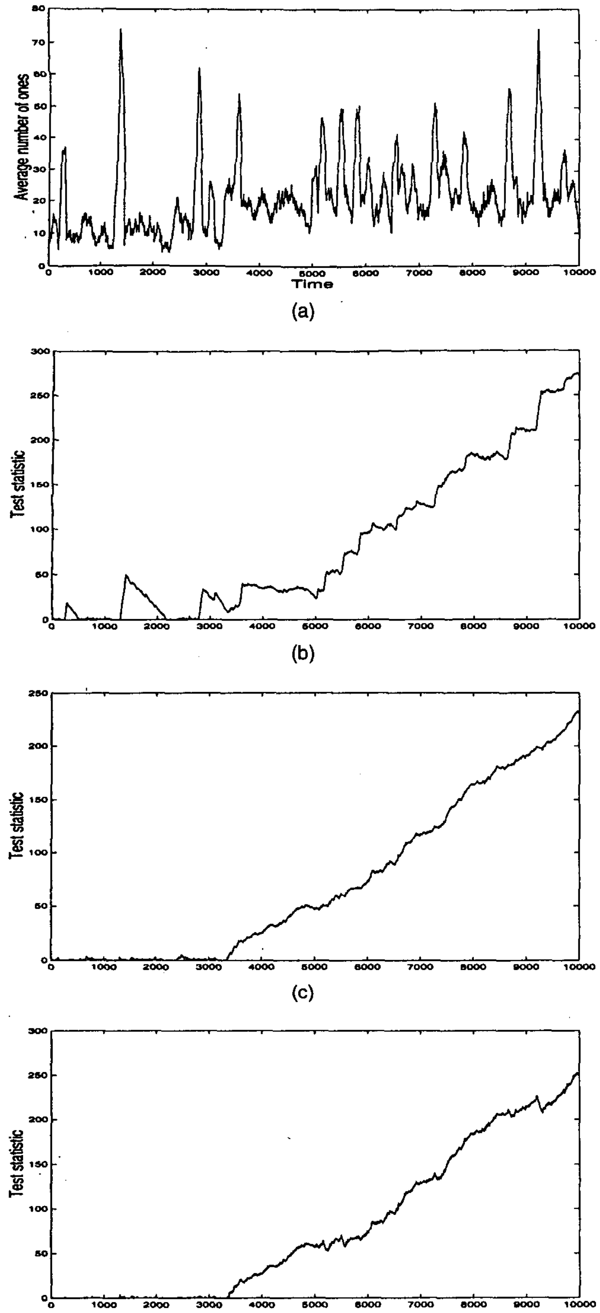


Fig. 4. Quickest detection of a second superimposed binary HMM which starts at 3333. (a) Number of 1s in each block of length 100. (b) "Naive" test result using the IID approximation. (c) Optimal scheme using HMM expansion. (d) Tracking-based approach.

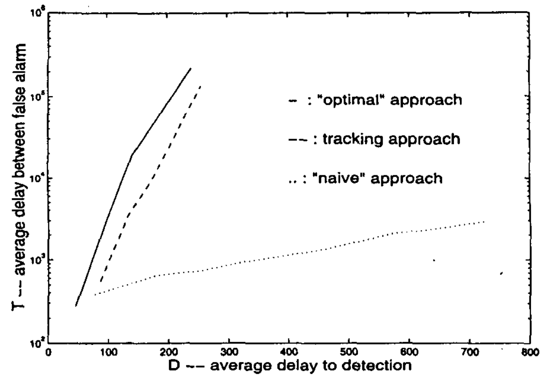


Fig. 5. ARL plot of the three schemes for the binary problem.

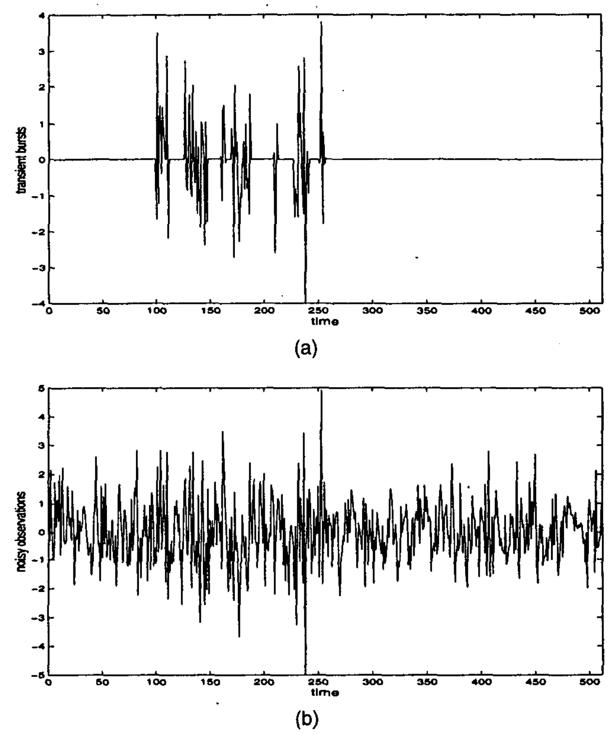


Fig. 6. Examples of Gaussian bursts transient signal. Transient starting time is 100. (a) Generated transient bursts. (b) Transient buried in Gaussian background.

matrix

$$A_1 = \begin{bmatrix} 0.95 & 0.05 \\ 0.10 & 0.90 \end{bmatrix}$$

One of the states is zero mean unity variance Gaussian while the other state is zero mean with variance 9. At an unknown instant, a second HMM similar to the first except for a different transition matrix

$$A_2 = \begin{bmatrix} 0.70 & 0.30 \\ 0.30 & 0.70 \end{bmatrix}$$

is active and added to the first HMM. In a sense, our procedure will test for different "clumps" of increased-variance observations, see Fig. 7. The ARL of the "optimal" and the tracking-based schemes are obtained via simulation and are plotted in Fig. 8. It is seen that for this specific problem, the "optimal" scheme has some advantage in terms of ARL, although both work reasonably well.

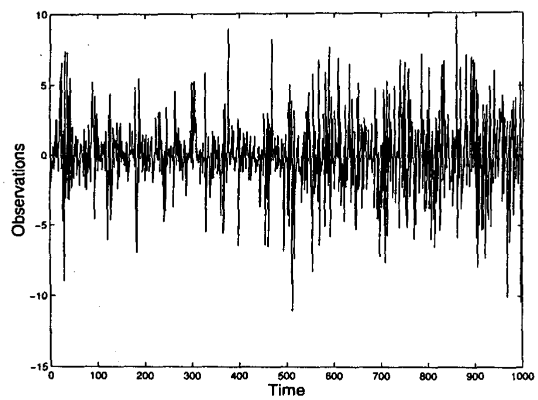


Fig. 7. Superimposed Gaussian increased-variance bursts HMMs. Second HMM starts at $n = 500$.

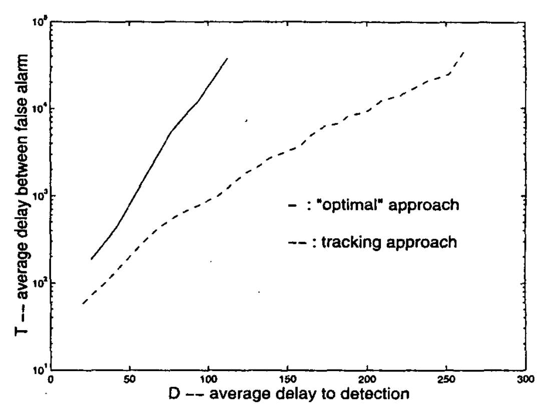


Fig. 8. ARL plot of the "optimal" and tracking approach for the Gaussian-shift-in-variance problem.

VI. CONCLUSIONS

HMMs provide a convenient and reasonably accurate characterization of certain signals of interest, particularly those of short duration and which undergo a cycle of behavior. Detection of such signals is important, and is complicated by the possibility of *superposition*: one or more extant HMM may obscure the onset of that of interest.

A technique has been reported elsewhere for detection of a switch from one HMM to another, and here we show that this idea is easily extended to multiple signals. The cost, of course, is numerical load. Target tracking is a special and heavily studied case of HMM state estimation. The goals (location versus likelihood evaluation) are different, but the ideas are similar. Thus borrowing from the target-tracking literature we develop a suboptimal JPDAF-like procedure for detection of superimposed HMMs. Its performance is reasonable, and its computational load is relatively light.

REFERENCES

[1] Bansal, R., and Papantoni-Kazakos, P. (1986)
An algorithm for detecting a change in a stochastic process.
IEEE Transactions on Information Theory, **32**, (Mar. 1986), 227–235.

[2] Baggenstoss, P., and Greineder, S. (1996)
Private communication.
NUWC Newport.

[3] Bar-Shalom, Y., and Li, X. (1995)
Multitarget-Multisensor Tracking: Principles and Techniques.
Storrs, CT: YBS Publishing, 1995.

[4] Basseville, M., and Nikiforov, L. (1993)
Detection of Abrupt Changes,
Englewood Cliffs, NJ: Prentice Hall, 1993.

[5] Bethel, R., and Paras, G. (1994)
A PDF multitarget tracker.
IEEE Transactions on Aerospace and Electronic Systems, **30**, (Apr. 1994), 386–403.

[6] Blackman, S., and Popoli, R. (1999)
Design and Analysis of Modern Tracking Systems,
Dedham, MA: Artech House, 1999.

[7] Chen, B., and Willett, P. (2000)
Quickest detection of hidden Markov models transient signals.
IEEE Transactions on Aerospace and Electronic Systems, **36**, 4 (Oct. 2000), 1253–1268.

[8] Dey, S., and Marcus, S. (1999)
Change detection in Markov-modulated time series.
Proceedings of Information, Decision, and Control, Feb. 1999.

[9] Fischer, W., and Meier-Hellstern, K. (1992)
The Markov-modulated Poisson process.
Performance Evaluation, **18**, (1992).

[10] Lorden, G., (1971)
Procedures for reacting to a change in distribution.
Annals of Mathematical Statistics, **42**, (June 1971), 1897–1908.

[11] Moustakides, G. (1986)
Optimal stopping times for detecting changes in distributions.
Annals of Statistics, **14**, (June 1986), 1379–1387.

[12] Page, E. (1954)
Continuous inspection schemes.
Biometrika, **41**, (1954), 100–115.

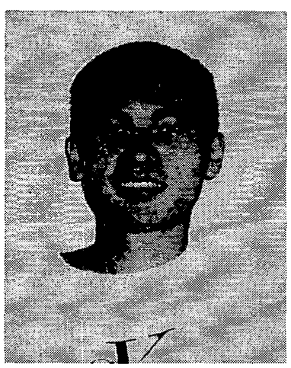
[13] Rabiner, L., and Juang, B. (1986)
An introduction to hidden Markov models.
IEEE ASSP Magazine, (Jan. 1986), 4–16.

[14] Siegmund, D. (1995)
Sequential Analysis—Tests and Confidence Intervals.
New York: Springer-Verlag, 1995.

[15] Streit, R., and Barrett, R. (1990)
Frequency line tracking using hidden Markov models.
IEEE Transactions on Acoustics, Speech, Signal Processing, **38**, (Apr. 1990), 586–598.

[16] Wald, A. (1947)
Sequential Analysis.
New York: Wiley, 1947.

[17] Wald, A., and Wolfowitz, J. (1948)
Optimum character of the sequential probability ratio test.
Annals of Mathematical Statistics, **19**, (1948), 326–339.



Biao Chen received his B.E. in 1992, M.A. in 1994, both in electrical engineering, from Tsinghua University, China.

He worked at AT&T, Inc., Beijing, during 1994–1995. He went to the University of Connecticut, Storrs, in 1995 where he obtained his Masters in statistics and Ph.D. in electrical engineering, in 1998 and 1999, respectively. After graduation, he spent a year at Cornell University as a Post-Doctoral Associate. He joined Syracuse University, Syracuse, NY, in 2000 as an Assistant Professor. His research area is in communications and signal processing, with particular interest in wireless communications.

Peter Willett (S'83—M'86) received the B.A.Sc. in 1982 from the University of Toronto, Toronto, Canada, and the Ph.D. from Princeton University, Princeton, NJ, in 1986.

He is as Associate Professor at the University of Connecticut, Storrs, where he has worked since 1986. His interests are generally in the areas of detection theory and signal processing, and, lately, particularly in the area of data fusion.