Correspondence

On the Inclusion of Channel's Time Dependence in a Hidden Markov Model for Blind Channel Estimation

Carles Antón-Haro, José A. R. Fonollosa, Claudi Faulí, and Javier R. Fonollosa

***Abstract*—In this paper, the theory of hidden Markov models (HMM) is applied to the problem of blind (without training sequences) channel estimation and data detection. Within a HMM framework, the Baum–Welch (BW) identification algorithm is frequently used to find out maximum-likelihood (ML) estimates of the corresponding model. However, such a procedure assumes the model (i.e., the channel response) to be static throughout the observation sequence. By means of introducing a parametric model for time-varying channel responses, a version of the algorithm, which is more appropriate for mobile channels [time-dependent Baum-Welch (TDBW)] is derived. Aiming to compare algorithm behavior, a set of computer simulations for a GSM scenario is provided. Results indicate that, in comparison to other Baum–Welch (BW) versions of the algorithm, the TDBW approach attains a remarkable enhancement in performance. For that purpose, only a moderate increase in computational complexity is needed.**

***Index Terms*—Blind channel estimation, GSM, hidden Markov models, mobile communications.**

I. INTRODUCTION

Whenever a signal propagates through a communications channel, several phenomena such as intersymbol interference, cochannel interference, and deep fades degrade its quality. In order to compensate for those effects and, eventually, estimate the transmitted data sequence, the use of receiver equalizers is mandatory. For the initial adjustment of such devices, conventional receivers resort to some sort of reference signal: training sequences, pilot symbols, or pilot tones. As soon as the start up period is over, further adjustment can be made in a decision-directed operation mode; that is, employing detected symbols as a reference signal.

In a communications scenario, making use of training sequences or any other type of side information reverts in a less efficient utilization of the assigned bandwidth. Moreover, in some applications (deconvolution of seismic traces, image, or recording restoration), resorting to a reference signal is not only barely recommended, but cumbersome or impossible. In consequence, the development of signal processing techniques which allow to deconvolve the received sequence (or, alternatively, estimate the channel impulse response) blindly (that is, with no knowledge on the transmitted data), has received great attention. Unfortunately, the blind deconvolution problem cannot be solved with traditional methods involving second-order moments of stationary scalar processes [1] so that blind approaches must consider a different set of hypothesis. For example, higher than second-order moments of the received signal can be taken into account either implicitly (Busggang methods [2]–[4]) or explicitly (higher order statistics-based approaches [5]–[7]). Other techniques aim to introduce some sort of spatial or temporal diversity in the received signal in order to perform blind estimation tasks on the basis of second-order moments of cyclostationary or vector processes [8]–[10]. Alternatively, the blind detection problem can be posed in terms of a joint identification process of both the channel characteristics and the transmitted data. In consequence, a maximum-likelihood (ML) approach, i.e., the maximization of a certain probability density function of the received data with respect to the unknowns, can be applied. This strategy is adopted by the so-called probabilistic algorithms [11]–[16], which, far from being restricted to second-, third-, or fourth-order moments, take into account all the statistical information embedded in the received signal.

This paper is aimed to develop probabilistic algorithms for blind channel identification on the basis of theory of hidden Markov models (HMM) [17]. HMM-based methods rely on the use of the Baum–Welch (BW) reestimation procedure [18], which is a particularization of the well-known expectation–maximization iterative algorithm [19]. By means of introducing a parametric model for time-varying channel responses, a version of the algorithm which is more appropriate for mobile channels [time-dependent Baum-Welch (TDBW)] is derived.

Adopting such a stochastic modeling for the received signal allows exploiting previous references in the literature, in particular, in the field of speech recognition. Nevertheless, HMM-based techniques are still far from being extensively used in digital communications problems.

II. SIGNAL MODEL

In the GSM system, a partial-response Gaussian minimum shift keying (GMSK) modulation scheme with an equivalent bandwidth of $\mathrm{BT} = 0.3$ is used. By making use of a partial response modulator (i.e., the modulator itself introduces intersymbol interference), the transmitted signal is granted with better spectral properties. At the output of the GSMK modulator, the low-pass equivalent for such a signal is given by the following expression (see Fig. 1):

$$d(t) = e^{j\phi(t,\mathbf{a})} = e^{j\pi \int_{-\infty}^{t} \sum_n a[n] g(\tau - nT)\, d\tau} \tag{1}$$

where

$g(t)$	stands for the phase-shaping pulse;
$a[n] \in \{1, -1\}$	are the transmitted symbols;
T	denotes symbol period.

For simplicity, a linear approximation to this phase modulation scheme will be derived. As shown in [20], any continuous-phase modulation (CPM) scheme can be decomposed in a finite sum of amplitude modulation terms. In particular, the transmitted signal in a GSM context can be accurately represented by

$$d(t) \approx \sum_n \alpha[n] j^n p(t - nT) \tag{2}$$

with $p(t)$ being a partial-response amplitude pulse and where the sequence $\alpha[n] \in \{1, -1\}$ is obtained by differentially encoding the se-

Manuscript received December 1, 1995; revised June 24, 1999. This work was supported in part by the National Plan of Spain (CICYT) TIC96-0500-C10-01, TIC98-0412, TIC98-0703, the Generalitat de Catalunya,(CIRIT) 1998SGR-00 081, and the European Commission under the ACTS AC347 SUNBEAM project.

C. Antón-Haro and J. R. Fonollosa are with the Department of Signal Theory and Communications, Universitat Politécnica de Catalunya, Barcelona, Spain (e-mail: carles@gps.tsc.upc.es; fonog@gps.tsc.upc.es).

J. A. R. Fonollosa and C. Faulí are with the Department of Signal Theory and Communications, Universitat Politécnica de Catalunya, Barcelona, Spain (e-mail: adrian@gps.tsc.upc.es).

Publisher Item Identifier S 0018-9545(01)04240-2.

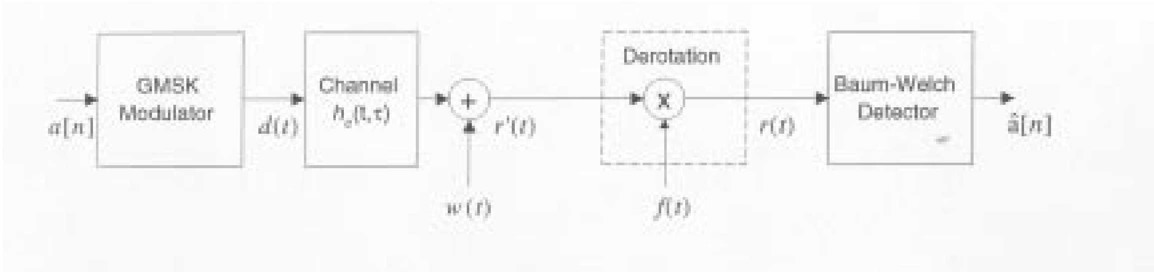


Fig. 1. Transmission subsystem.

quence of transmitted symbols ($\alpha[n] = [n-1]a[n]$). The received signal can be expressed in the following terms:

$$r'(t) = d(t) * h_c(t,\tau) \approx \sum_n \alpha[n] j^n h'(t-nT,t) \tag{3}$$

where $h'(t,\tau) = h_c(t,\tau) * p(t)$ stands for the overall channel response including transmit $(p(t))$ and receive filters and the physical channel response itself $(h_c(t,\tau))$. Further, signal goes through a *derotation* stage

$$\begin{aligned} r(t) = r'(t)f(t) &= \sum_n \alpha[n] j^n (-j)^n f(t-nT) h'(t-nT,t) \\ &= \sum_n \alpha[n] h(t-nT,t) \end{aligned} \tag{4}$$

where $h(t,\tau) = h'(t,\tau)f(t) = h(t,\tau)\sum_n (-j)^n \Pi(t/T)$. In summary, after derotation the CPM-modulated received signal can be viewed as a pulse-amplitude modulation (PAM)-modulated one being corrupted by an equivalent linear channel spanning along $L = L_c + L_m - 1$ symbol periods. In the latter expression, L_c accounts for the amount of ISI introduced by the physical channel whereas L_m reflects the influence of the partial impulse response modulator. Hereinafter, we will assume the physical channel memory to be increased in two symbol periods; that is, $L = L_c + 2$.

Taking into account the contribution of the additive noise $w(t)$ and sampling the received signal at the symbol rate, an equivalent vector expression for the received signal is yielded

$$r[n] = h[n]^T \mathbf{s}[n] + w[n] \tag{5}$$

where $\mathbf{h}[n] = [h(nT,nT), h((n-1)T,nT), \ldots, h((n-L+1)T,nT]^T$. In other words, the sequence of observations $\{r[n]\}$ can be modeled as a probabilistic function of the state vector $\mathbf{s}[n] = [\alpha[n], \alpha[n-1], \ldots, \alpha[n-L+1]]^T$. Since, at any given time, a maximum of L symbols affect the observation, there are $N_s = 2^L$ possible state vectors corresponding to all combinations of L binary symbols. We will denote each of the possible states as the L-length vector $\mathbf{s}_j = [s_j^{(0)}, s_j^{(1)} \ldots, s_j^{(L-1)}]^T$ with $s_j^{(l)} \in \{1,-1\}$ and the actual state at time instant nT as $\mathbf{s}[n] \in \mathbf{S}$. Yet, we will define the states matrix

$$\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{N_s}]^T \tag{6}$$

where each row contains the symbols for every state in the model.

III. Overview of the BW Algorithm

The signal model introduced in the previous section can be used to describe the received sequence in terms of a HMM [18]. An HMM is a doubly stochastic process with an underlying stochastic process that is not observable (hidden), but can only be observed through another set of stochastic processes that produce the sequence of observed symbols. If a set of statistical parameters—the *model*—is obtainable, it can then be used to identify or recognize other sequences of observations. In the case we are considering, transmitted data constitute the hidden process, whereas the received sequence plays the role of the observable process. Such processes can be characterized by the following parameter set.

1) The number of states: $N_s = 2^L$; that is, the number of distinct inputs the system may have for a given observation.
2) The state transition matrix:

$$\begin{aligned} \mathbf{A} &= \{a_{ij}\} \quad 1 \le i,\, j \le N_s \\ a_{ij} &= \Pr(\mathbf{s}[n+1] = \mathbf{s}_j \,|\, \mathbf{s}[n] = \mathbf{s}_i) \\ &= \begin{cases} \Pr\left(s_j^{(0)}\right), & \text{if } s_j^{(l)} = s_j^{(l-1)} \quad l = 0 \ldots L-1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{7}$$

3) The vector of probability density functions of the observation $r[n]$ conditioned on a given state channel $\mathbf{s}_j$:

$$\mathbf{B} = [b_1(r[n]), \ldots, b_{N_s}(r[n])]^T. \tag{8}$$

In the presence of AWGN noise, each element in the array amounts to

$$\begin{aligned} b_j(r[n]) &= p(r[n] \,|\, \mathbf{s}_j) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(r[n]-m_j)^H (r[n]-m_j)}{2\sigma^2}\right) \end{aligned} \tag{9}$$

with $m_j = \mathbf{h}^T \mathbf{s}_j$, $1 \le j \le N_s$.

The initial state distribution vector $\pi = [\pi_1, \ldots, \pi_{N_s}]^T$ where $\pi_i = \Pr(\mathbf{s}[0] = \mathbf{s}_i)$ is assigned an arbitrary value, say $\pi_i = 1/N_s$. In the sequel, we will refer to such a HMM using the short-hand notation $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$. In this paper, we will assume that L is either known or can be upper bounded and that the statistics of the transmitter symbols are also known. Thus, by exclusively keeping the unknowns in the model and defining $\mathbf{m} = [m_1, m_2, \ldots, m_{N_s}]^T$, the model can be represented by the parameter set $\lambda = (\mathbf{m}, \sigma^2)$ or, equivalently, $\lambda = (\mathbf{h}, \sigma^2)$.

The BW algorithm is an iterative procedure[1] which is known to provide an ML solution for the parameter set λ. It can be shown [17] that for a given set of N observations, the average log-likelihood function to be maximized amounts to

$$\begin{aligned} Q\left(\lambda, \lambda^{(k)}\right) = \mathrm{C} &+ \sum_{n=1}^{N} \sum_{i=1}^{N_s} \gamma_i^{(k)}[n] \\ &\times \left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|r[n]-m_i\|^2\right) \end{aligned} \tag{10}$$

[1]Actually, a particularization of the well-known expectation-maximization algorithm [19]).

where $\gamma_i^{(k)}[n]$ stands for the probability of being in state $\mathbf{s}_i$ at time instant n, given the whole sequence of observations and the estimated model after the kth iteration.[2] Thus, in order to maximize $Q(\lambda, \lambda^{(k)})$, the following recursion should be applied:

$$\hat{\mathbf{m}}^{(k+1)} = \arg\max_{\mathbf{m}} Q\left(\lambda, \lambda^{(k)}\right) \tag{11}$$

$$\hat{\sigma}^{2(k+1)} = \arg\max_{\sigma^2} Q\left(\lambda, \lambda^{(k)}\right) \tag{12}$$

thus providing a new estimate of the model $\lambda^{(k+1)}$. More precisely

$$\nabla_{\mathbf{m}} Q(\lambda, \lambda^{(k)}) = 0 \Rightarrow \hat{m}_i^{(k+1)} = \frac{\sum_{n=1}^{N} \gamma_i^{(k)}[n] r[n]}{\sum_{n=1}^{N} \gamma_i^{(k)}[n]} \tag{13}$$

$$\nabla_{\sigma^2} Q(\lambda, \lambda^{(k)}) = 0 \Rightarrow \hat{\sigma}^{2(k+1)} = \frac{\sum_{n=1}^{N} \sum_{i=1}^{N_s} \gamma_i^{(k)}[n] \left\| r[n] - \hat{m}_i^{(k+1)} \right\|^2}{\sum_{n=1}^{N} \sum_{i=1}^{N_s} \gamma_i^{(k)}[n]}. \tag{14}$$

This procedure is iterated until an application-specific convergence criterion is fulfilled. At that time, data detection can be performed following, for example, a *maximum a posteriori* (MAP) criterion. At instant n, the index to the MAP state is

$$\mathrm{i}_n = \arg\max_{1 \le i \le N_s} [\gamma_i[n]] \quad 1 \le n \le N \tag{15}$$

where $\gamma_i[n]$ is the final value for $\gamma_i^{(k)}[n]$. Accordingly, an estimate of the transmitted data symbols can be obtained by properly selecting among the element of the MAP state.

To point out that when the channel can be characterized in terms of a finite impulse response (FIR) filter, an improved estimate of the channel response can be obtained after every reestimation step. To do so, a least squares (LS) projection is used

$$\hat{\mathbf{m}}^{(k)} \leftarrow \mathcal{S}\mathcal{S}^{\#} \hat{\mathbf{m}}^{(k)} = \mathbf{P}_{\mathcal{S}} \hat{\mathbf{m}}^{(k)} \tag{16}$$

where $\mathbf{S}^{\#} = (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H$ denotes pseudoinverse. Thus, an LS estimate for the channel coefficients can be obtained from $\mathbf{m}$:

$$\mathbf{h}^{(k)} = \mathcal{S}^{\#} \mathbf{m}^{(k)}. \tag{17}$$

In the sequel, this version of the BW algorithm will be referred to as BW algorithm with LS projection step (BW-LS). Other adaptive versions of the algorithm [adaptive Baum–Welch (ABW)] relying on an LMS-like update scheme exist [16], [13].

[2]Variable $\gamma_i^{(k)}[n]$ can be efficiently computed making use of the forward–backward algorithm (see [17] for details).

IV. Modified Algorithm

In the generic (batch) version of the BW algorithm, the channel is assumed to be stationary during the data burst. This assumption, however, could not be appropriate when high-speed mobile stations are considered. On the other hand, adaptive versions (ABW) are able to perform channel tracking, but at the expense of robustness against fading. Actually, performance is severely impaired when a deep fade turns up during a signal burst.

Aiming to gather the advantages from both approaches in a new single scheme, the TDBW algorithm is proposed. Within the TDBW framework, the evolution of every tap in the vector channel response is approximated by means of a polynomial in n

$$\hat{\mathbf{h}}[n] = \mathbf{h}_0 + \mathbf{h}_1 \cdot n + \mathbf{h}_2 \cdot n^2 + \cdots. \tag{18}$$

Accordingly, the following parametric model is obtained for the time-varying means vector:

$$\begin{aligned} \hat{\mathbf{m}}[n] &= \mathbf{S}\hat{\mathbf{h}}[n] && (19)\\ &= \mathbf{S}(\mathbf{h}_0 + \mathbf{h}_1 \cdot n + \mathbf{h}_2 \cdot n^2 + \cdots) && (20)\\ &= \mathbf{m}_0 + \mathbf{m}_1 \cdot n + \mathbf{m}_2 \cdot n^2 + \cdots. && (21) \end{aligned}$$

Next, each element $m_i[n]$ in this latter vector expression will be included in the cost function in (10)

$$\begin{aligned} Q\left(\lambda, \lambda^{(k)}\right) = \mathrm{C} &+ \sum_{n=1}^{N} \sum_{i=1}^{N_s} \gamma_i^{(k)}[n] \\ &\times \left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \| r[n] - m_{i,0} \right. \\ &\left. - \, m_{i,1} \cdot n - m_{i,2} \cdot n^2 \ldots \|^2 \right). \end{aligned} \tag{22}$$

By doing so, the *time-varying* nature of the channel response is embedded in a set of *batch* BW formulas. In order to derive an expression to reestimate the parameters in the model, we will restrict to first- and second-order approximations. For moderate speed users (less than 100 km/h), the first-order approximation

$$\hat{\mathbf{h}}_j = \mathbf{h}_0 + \mathbf{h}_1 \cdot n \tag{23}$$

was observed to be accurate enough. Again, taking partial derivatives with respect to m_0 and m_1 leads to (24)–(26), shown at the bottom of the page, which provide a new estimate for the components of the means vector and the variance of the additive noise. The following definitions apply:

$$A = \sum_{n=1}^{L} \gamma_i^{(k)}[n] \tag{27}$$

$$m_{i,0}^{(k+1)} = \frac{A \cdot \left(\sum_{n=1}^{L} \gamma_i^{(k)}[n] r[n] \cdot n\right) - B \cdot \left(\sum_{n=1}^{L} \gamma_i^{(k)}[n] r[n]\right)}{\Delta_2} \tag{24}$$

$$m_{i,1}^{(k+1)} = \frac{C \cdot \left(\sum_{n=1}^{L} \gamma_i^{(k)}[n] r[n]\right) - B \cdot \left(\sum_{n=1}^{L} \gamma_i^{(k)}[n] r[n] \cdot n\right)}{\Delta_2} \tag{25}$$

$$\hat{\sigma}^{2(k+1)} = \frac{1}{L} \sum_{n=1}^{L} \sum_{i=1}^{N} \gamma_i^{(k)}[n] \left\| r[n] - m_{i,0}^{(k+1)} - m_{i,1}^{(k+1)} \cdot n \right\|^2 \tag{26}$$

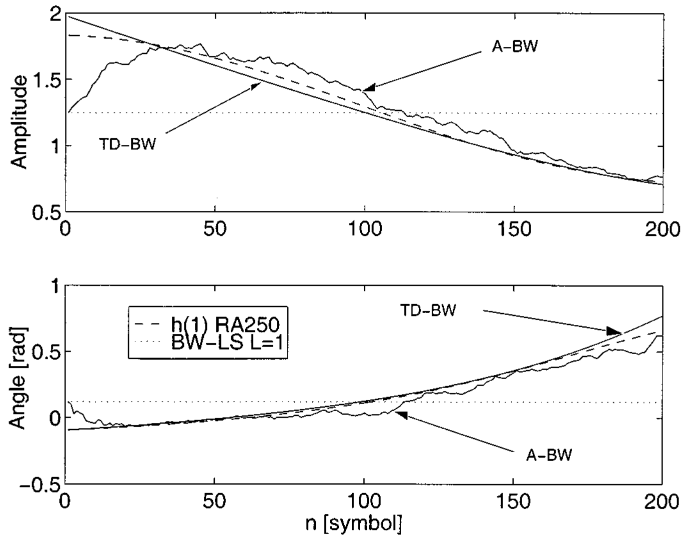


Fig. 2. Tracking for the first tap in a RA250 test channel using a first-order TDBW approximation: (a) rectangular coordinates and (b) magnitude and angle. Dashed lines: actual evolution of the tap, dotted line: BW-LS estimate.

$$B = \sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot n \tag{28}$$

$$C = \sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot n^2. \tag{29}$$

This result constitutes a local maximum of the log-likelihood function (see Appendix) Fig. 2 reflects the ability of the TDBW and ABW approaches to track the first tap in a Rural Area channel ($v = 250$ km/h). It should be noted that, by far, the TDBW estimate is less noisy than that of the ABW version (adaptation step: $\mu = 0.05$). For comparison, the static channel estimate provided by the BW-LS algorithm is also depicted (dotted line).

For high-speed mobiles, though, we can do better by increasing polynomial order. Making use of a second-order approximation ($\hat{\mathbf{m}} = \mathbf{m}_0 + \mathbf{m}_1 \cdot n + \mathbf{m}_2 \cdot n^2$) along with the procedure described in (30)–(33), shown at the bottom of the page, can be obtained where

$$f = \sum_{n=1}^{L} \gamma_i^{(k)}[n] r[n]; \quad g = \sum_{n=1}^{L} \gamma_i^{(k)}[n] r[n] \cdot n;$$

$$h = \sum_{n=1}^{L} \gamma_i^{(k)}[n] r[n] \cdot n^2$$

$$D = \sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot n^3; \quad E = \sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot n^4$$

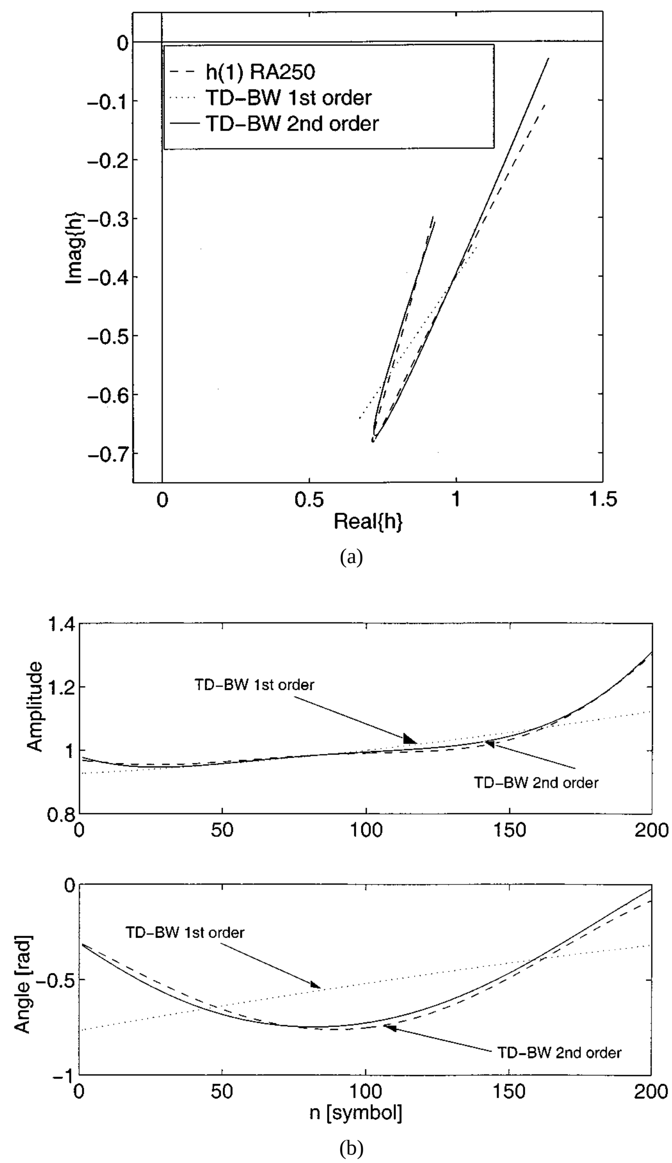


Fig. 3. Tracking for the first tap in a RA250 test channel using first- and second-order TDBW approximations: (a) rectangular coordinates and (b) magnitude and angle. Dashed lines: actual evolution of the tap.

$$\Delta_2 = \det \begin{pmatrix} A & B & C \\ B & C & D \\ C & D & E \end{pmatrix} \tag{34}$$

and A, B, and C were defined previously. First- and second-order approximations are compared for a specific channel realization in Fig. 3.

$$\hat{m}_{i,0}^{(k+1)} = \frac{(CE - D^2) \cdot f - (BE - CD) \cdot g + (BD - C^2) \cdot h}{\Delta_2} \tag{30}$$

$$\hat{m}_{i,1}^{(k+1)} = \frac{-(BE - DC) \cdot f - (AE - C^2) \cdot g + (AD - CB) \cdot h}{\Delta_2} \tag{31}$$

$$\hat{m}_{i,2}^{(k+1)} = \frac{(BD - C^2) \cdot f - (AD - BC) \cdot g + (AC - B^2) \cdot h}{\Delta_2} \tag{32}$$

$$\hat{\sigma}^{2(k+1)} = \frac{1}{L} \sum_{n=1}^{L} \sum_{i=1}^{N} \gamma_i^{(k)}[n] \left\| r[n] - m_{i,0}^{(k+1)} - m_{i,1}^{(k+1)} \cdot n - m_{i,2}^{(k+1)} \cdot n^2 \right\|^2 \tag{33}$$

Clearly, a more accurate estimate is obtained when the second order approximation is used. Nevertheless, performance gain should be evaluated in terms of BER versus $E_b = N_o$. This will be done in the next section.

A. Discussion

As far as computational complexity is concerned, it becomes apparent that the need to estimate more parameters in the TDBW versions than in the original BW and BW-LS approaches ($\mathbf{m}_0$, $\mathbf{m}_1$ and $\mathbf{m}_2$ versus $\mathbf{m}$) reverts in an increased computational burden. This additional complexity, though, is much lower than that of computing variable $\gamma_i^{(k)}[n]$, which, as shown in [17], depends on the square of the number of states, i.e., $O(N_s^2)$. In consequence, computational complexity *per pass* for the TDBW algorithm is not that different from those of the previous versions.

However, it should be noted that in the second-order TDBW version three means vectors (instead of one) must be estimated from the same number of observations (N_s, one data burst). For reduced data sets where $N \approx N_s$, this could jeopardize the resulting estimates for the means since some states could be seldom visited or, even, not visited at all. Aiming to circumvent this difficulty, the following strategy was adopted: replacing the LS channel estimation by a weighted LS channel estimation ($\mathbf{h}^{(k)} = (\mathbf{S}^H\mathbf{W}^{(k)}\mathbf{S})^H\mathbf{S}^H\mathbf{W}^{(k)}\mathbf{m}^{(k)}$). This way, despite that for some states the means estimates could still be unsatisfactory, those errors will not propagate to the next iteration step. The optimal choice for $\mathbf{W}$ is the inverse of the error correlation matrix. However, such matrix is difficult to obtain and, hence, a heuristic approximation will be used instead. More precisely, assuming that estimation errors are independent among states, a diagonal weighting matrix will be utilized

$$\mathbf{W}^{(k)} = k_w \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & w_N \end{pmatrix} \tag{35}$$

where

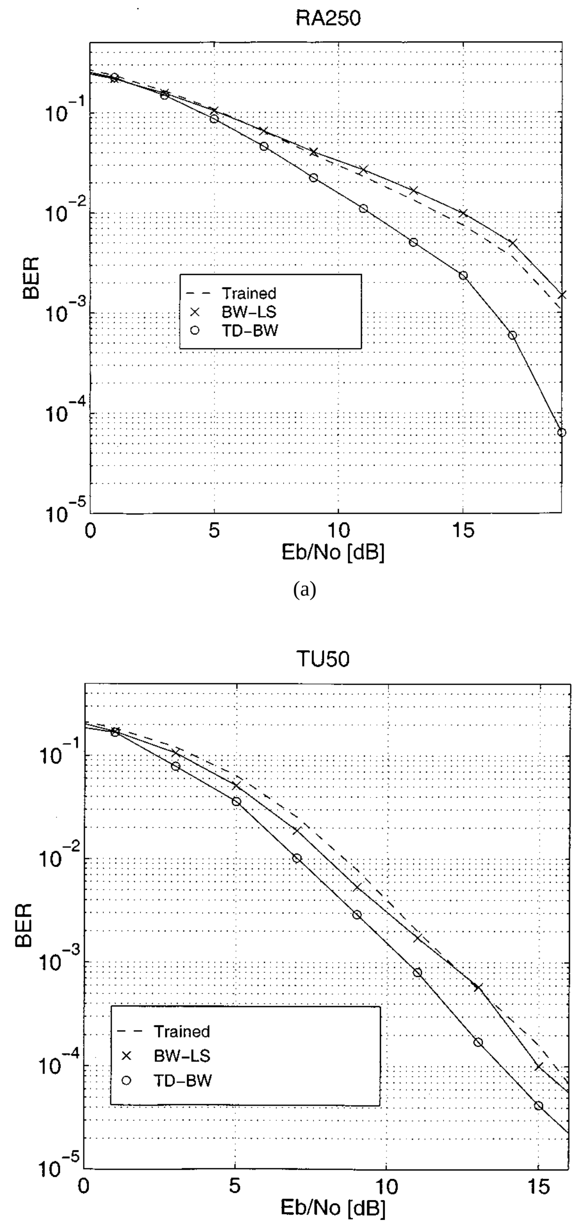
$$w_i^{(k)} = \sum_{n=1}^{L} \gamma_i^{(k)}[n] \tag{36}$$

and k_w is an arbitrary scale factor. Matrix elements $w_i^{(k)}$ provide a measure of the reliability in the estimation of every component of vector $\mathbf{m}$. Actually, it can be shown that such sums are *a posteriori* estimates of the number of times that every state was observed over the sequence of observations.

V. Simulation Results

Algorithm performance was assessed for the standard test channels described in the ETSI recommendations [21]. On the other hand, mobiles' velocity in each environment was chosen according to [22]. With no doubt, the most interesting case were RA250 and TU50 (rural area environment-100 km/h, and typical urban 50 km/h) since they illustrate to what extent the TDBW version may be useful in each situation. As for channel memory, parameter L was set to three and four symbols for the RA and TU cases, respectively.

Fig. 4 depicts the BER versus instantaneous (in-burst) E_b/N_o curves for the BW-LS and TDBW approaches. A benchmark curve reflecting the behavior exhibited by a nonblind approach (channel estimation using the mid-amble training sequence plus MLSE detector) is provided as well. Also, a plot showing error distributions along the signal burst is included (Fig. 5).

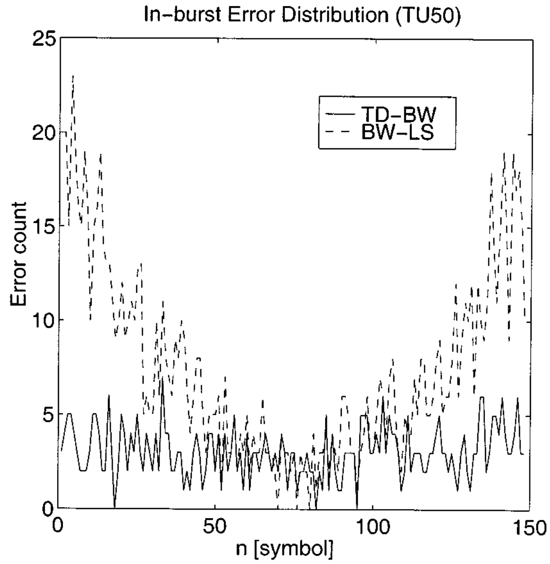


(a) (b)

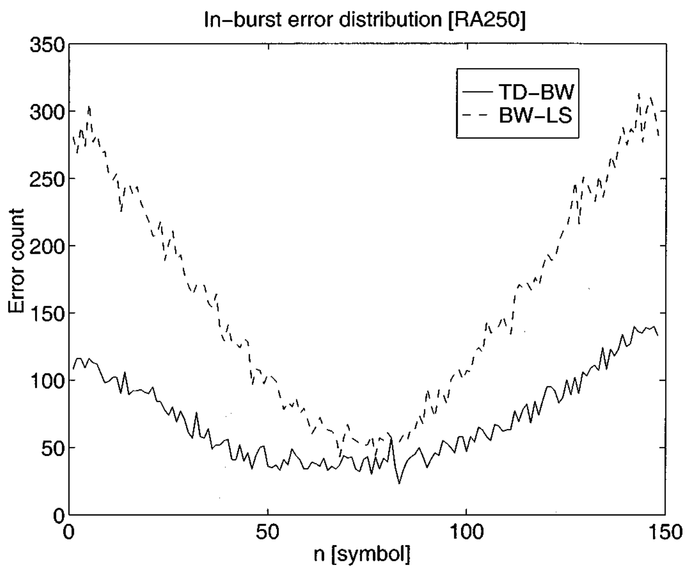
Fig. 4. Bit error rate versus E_b/N_o. (a) Test channel: RA250. (b) TU50.

In both cases, performance exhibited by the BW-LS version and the benchmark receiver are very close. On the other hand, the first-order TDBW version clearly outperforms both of them. This is particularly true for the RA250 test environment where the increased vehicle speed reduces channel coherence time-intervals. As shown in the error histograms, such an improvement is motivated by the enhanced ability to track channel variations around burst ends. As a matter of fact, it is in those regions where the static channel estimates differ to a greater extent from the actual channel taps (see Fig. 2).

Last, as far as convergence criterion is concerned, it should be noted first that both the actual variances of the additive noise and the channel impulse response are unknown. Consequently, the selected convergence criterion cannot be expressed in terms of euclidean distances to the actual model. Instead, the burst-averaged relative distance between successive channel estimates was used. By observing that, as soon as the MSE between channel estimates falls below $5 \cdot 10^{-4}$ BER curves do not differ significantly, a threshold for such a distance was



(a)



(b)

Fig. 5. In-burst error distribution averaged over 5000 independent runs at $E_b/N_o = 20$ dB. (a) Test channel: RA250. (b) TU50.

empirically determined. Fig. 6 illustrates the number of iterations needed for the algorithms to converge versus E_b/N_o. As observed, at $E_b/N_o = 9$ dB ten iterations are needed when using the BW-LS approach and 15 when using TDBW. This increase is motivated by the fact that in order to properly initialize the TDBW scheme, an static BW-LS channel estimate had to be computed first. Roughly, the ratio between blind BW-based methods and trained Viterbi-based receivers' complexity is given by the number of iterations up to convergence.

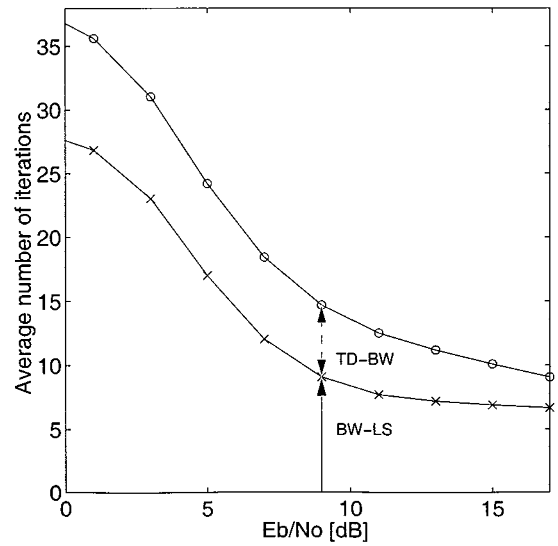


Fig. 6. Average number of iterations to convergence versus E_b/N_o for the TDBW and BW-LS algorithms.

VI. Conclusion

A new version of the BW estimation procedure has been presented and its performance assessed in a GSM scenario. Also, a mechanism relying on the use of an heuristic (albeit effective) weighting matrix is introduced in order to cope with reduced data sets. Results indicate that performance can be remarkably improved by including the time-varying nature of the channel response in the batch reestimation formulae. This can be done with a moderate increase in computational burden with respect to the BW-LS version. For realistic E_b/N_o values, though, computational complexity for both blind approaches is significantly higher than those of trained methods.

Appendix
Proof: Maximization of the Log-Likelihood Function

Right after the k(th) iteration, the Hessian of the log-likelihood function [see (37) at the bottom of the page] is negative definite as long as the determinants

$$\Delta_1 = A = \sum_{n=1}^{L} \gamma_i^{(k)}[n] \tag{38}$$

$$\Delta_2 = AC - B^2 = \sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot \sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot n^2 - \left(\sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot n \right)^2 \tag{39}$$

are both strictly positive. Hence, this result constitutes a local maximum of the cost function unless

$$\sum_{n=1}^{L} \gamma_i^{(k)}[n] = 0 \tag{40}$$

$$\mathbf{H}^{(k)} = -2 \cdot \begin{pmatrix} \sum_{n=1}^{L} \gamma_i^{(k)}[n] & \sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot n \\ \sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot n & \sum_{n=1}^{L} \gamma_i^{(k)}[n] \cdot n^2 \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} A & B \\ B & C \end{pmatrix} \tag{37}$$

(i.e., that state was not observed along the timeslot) or

$$\sum_{n=1}^{L} \gamma_i^{(k)}[n] = \gamma_i[n_i] \tag{41}$$

(i.e., that state was observed only once, for $n = n_i$). In those cases, of course, there is no point in looking for a linear approximation for the evolution of the channel response. □

REFERENCES

[1] S. Haykin, *Blind Deconvolution*. Englewood Cliffs, NJ: Prentice-Hall, 1994. Information and Science.

[2] Y. Sato, “A method for self-recovering equalization,” *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, June 1975.

[3] D. Godard, “Self-recovering equalization and carrier-tracking in two-dimensional data communication systems,” *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.

[4] S. Verdú, “Blind equalization without gain identification,” *IEEE Trans. Inform. Theory*, vol. 39, pp. 292–297, Jan. 1993.

[5] O. Shalvi and E. Weinstein, “New criteria for blind deconvolution of nonminimum phase systems,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 312–321, Mar. 1990.

[6] G. B. Giannakis, “Cumulants: A powerful tool in signal processing,” *Proc. IEEE*, pp. 1333–1334, 1987.

[7] J. M. Mendel, “Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications,” *Proc. IEEE*, vol. 79, pp. 277–305, Mar. 1991.

[8] W. A. Gardner, “Exploitation of spectral redundancy in cyclostationary signals,” *IEEE Signal Processing Mag.*, pp. 14–36, Apr. 1991.

[9] L. Tong, G. Xu, B. Hassibi, and T. Kailath, “Blind channel identification based on second-order statistics: A frequency domain approach,” *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Jan. 1995.

[10] K. Abed-Merain, J. F. Cardoso, A. Gorokhov, P. Loubaton, and E. Moulines, “On subspace methods for blind identification of single-input multiple-output multichannel FIR systems,” *IEEE Trans. Signal Processing—Special Issue: Signal Processing Adv. Commun.*, vol. 45, Jan. 1997.

[11] M. Ghosh and C. L. Weber, “Maximum-likelihood blind equalization,” in *Proc. SPIE Adaptive Signal Processing*, vol. 1565, 1991, pp. 188–195.

[12] V. Krishnamurthy and J. Moore, “On-line estimation of hidden Markov model parameters based on the Kullback–Leibler information measure,” *IEEE Trans. Signal Processing*, vol. 41, pp. 2557–2573, Aug. 1993.

[13] J. A. R. Fonollosa and J. Vidal, “Application of hidden Markov models to blind channel characterization and data detection,” *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 185–188, Apr. 1994.

[14] J. R. Fonollosa, J. A. R. Fonollosa, Z. Zvonar, and J. Vidal, “Blind multiuser identification and detection in CDMA systems,” *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 1876–1879, May 1995.

[15] C. Antón-Haro, J. A. R. Fonollosa, and J. R. Fonollosa, “Blind channel estimation and data detection using hidden Markov models,” *IEEE Trans. Signal Processing—Special Issue: Signal Processing Adv. Commun.*, vol. 45, no. 1, pp. 241–246, Jan. 1997.

[16] ——, “Probabilistic algorithms for blind adaptive multiuser detection,” *IEEE Trans. Signal Processing*, vol. 46, pp. 2953–2966, Nov. 1998.

[17] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77, pp. 257–286, Feb. 1989.

[18] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, “A maximization technique ocurring in the statistical analysis of probabilistic functions of Markov chains,” *Ann. Math. Statist.*, vol. 41, no. 1, pp. 164–171, 1970.

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. Roy. Statist. Soc.*, ser. 39, pp. 1–38, 1977.

[20] T. Aulin and C. W. Sundberg, “Continuous phase modulations (pts. I, II),” *IEEE Trans. Commun.*, vol. COM-29, pp. 196–225, Mar. 1981.

[21] “GSM Recommendation 05.05,” ETSI, Tech. Rep. Version 3.11.0, Mar. 1990. ETSI-GSM.

[22] R. d’Avella, L. Moreno, and M. Sant’Agostino, “An adaptive MLSE receiver for TDMA digital mobile radio,” *IEEE J. Select. Areas Commun.*, vol. 7, pp. 122–129, Jan. 1989.