Blind Channel Estimation and Data Detection Using Hidden Markov Models

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

***Abstract*— In this correspondence, we propose applying the hidden Markov models (HMM) theory to the problem of blind channel estimation and data detection. The Baum–Welch (BW) algorithm, which is able to estimate all the parameters of the model, is enriched by introducing some linear constraints emerging from a linear FIR hypothesis on the channel. Additionally, a version of the algorithm that is suitable for time-varying channels is also presented. Performance is analyzed in a GSM environment using standard test channels and is found to be close to that obtained with a nonblind receiver.**

I. Introduction

Blind equalization, i.e., the ability of initially adjust the equalizer in the receiver without training sequences or other side information, has received great attention in recent years. Blind equalization/estimation methods developed so far can be classified in three families:

1) Bussgang algorithms [1], [2]
2) polyspectra and cumulant-based algorithms [1], [3]
3) probabilistic algorithms. [4]–[10].

The algorithm proposed in this correspondence belongs to this third group. Probabilistic algorithms are based on optimal approaches that lead to joint channel estimation and data detection, often on a basis of a maximum likelihood (ML) criterion. These recently proposed methods exhibit higher computational complexity, but they clearly outperform Bussgang and polyspectra methods, for example, in terms of a more accurate channel identification from a very reduced number of samples [5]. Moreover, modeling the received signal as a hidden Markov model (HMM) allows us to exploit the rich literature in this field, particularly in speech recognition applications [11]. The HMM stochastic signal model (Markov sources or probabilistic functions of Markov chains) is not new in the communication literature, and Viterbi decoders may be derived from this formulation.

II. A Hidden Markov Model for the GSM System

Consider the transmission of a sequence independent symbols through an additive white Gaussian noise (AWGN) channel with finite memory. The received signal (Fig. 1) can be expressed as

$$x[n] = f(\mathbf{s}[n]) + w[n] \tag{1}$$

where $\mathbf{s}[n]$ is present state of the transmitter, and $w[n]$ is zero-mean AWGN with variance equal to σ^2. Observe that function $f(.)$ may include the effect of linear or nonlinear modulation schemes such as GMSK [12], but in order to model the signal as a HMM, finite-length memory is required to it. Function $f(.)$ is stated as

$$\begin{aligned} f(\mathbf{s}[n]) &= \sum_{i=0}^{L_c-1} h_i\, d[n-i] \\ &= \sum_{i=0}^{L_c-1} h_i\, e^{j\,\phi[n-i]} \end{aligned} \tag{2}$$

Manuscript received November 10, 1995; revised August 21, 1996. This work was supported by the Fundació Catalana per a la Recerca, the CIRIT-Generalitat de Catalunya (GRQ93-3021), and the CICYT of Spain (Grants TIC95-1022-C05-1 and TIC96-0500-C10-01).

The authors are with the Universitat Politècnica de Catalunya, Department of Signal Theory and Communications, 08034 Barcelona, Spain.

Publisher Item Identifier S 1053-587X(97)00528-X.

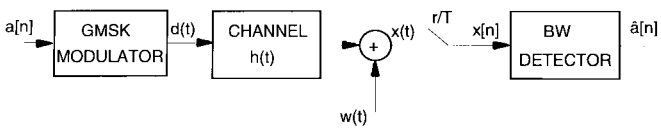

Fig. 1. Communications subsystem for the GSM standard coupled to a BW detector.

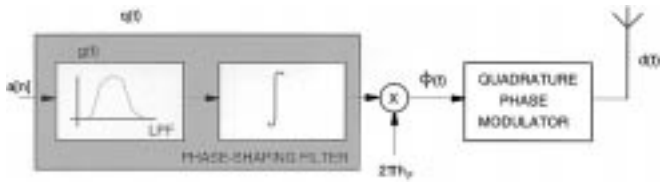

Fig. 2. Block diagram for a GMSK modulator.

where $d[n]$ is the discrete-time version of the GMSK signal. Such a signal $d(t)$ is the output of a quadrature phase modulator (Fig. 2) whose input is given by

$$\begin{aligned} \phi(t) &= 2\,\pi\,h_F \int_{-\infty}^{t} \sum_{m=-\infty}^{\infty} a[m]\, g(\tau - mT)\, d\tau \\ &= 2\,\pi\,h_F \sum_{m=-\infty}^{n} a[m]\, q(t-mT), \\ & \qquad nT \le t \le (n+1)T \end{aligned} \tag{3}$$

where $a[n] \in \{1, -1\}$. For a modulation index $h_F = 0.5$ and taking into account that the response of the phase shaping filter $q(t)$ equals to 0.5 for $t \ge L_m T$ (Fig. 3), we rewrite $\phi(t)$:

$$\begin{aligned} \phi(t) &= \pi \sum_{m=n-L_m+1}^{n} a[m]\, q(t-mT) + \pi\, h_F \sum_{m=-\infty}^{n-L_m} a[m] \\ &= \pi \sum_{m=n-L_m+1}^{n} a[m]\, q(t-mT) + \theta[n], \\ & \qquad nT \le t \le (n+1)T. \end{aligned} \tag{4}$$

As we can see, $\phi(t)$ depends on 1) the L_m most recent symbols and 2) $\theta[n] \in \{0, \pi/2, \pi, 3\pi/2\}$, namely, the accumulated phase coming from all the previous symbols that have completely passed through the filter up to instant n. Therefore, from (2) and (4), we conclude that the number of transmitter symbols (bits) involved in a single observation at the receiver is $L = L_m + L_c - 1$. However, for a relative bandwidth $BT = 0.3$ [12], the amount of ISI produced by the GMSK modulator can be neglected without significant performance loss.[1] At this point, we can model the observation $x[n]$ as a probabilistic function of the state $\mathbf{s}[n] = (a[n], \cdots, a[n-L+1], \theta[n])^T$ obtaining a description of the received sequence as a first-order HMM $\lambda = (\mathbf{A}, \mathbf{B}, \pi)$ [11] with the following characteristics:

1) The number of states is $N = 4 \cdot 2^L$, i.e., the number of distinct inputs that $f(.)$ may have. We denote the individual states as $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N]^T$ and the state at time n as $\mathbf{s}[n]$.

[1] Note that we are not assuming an approximately linear model for the GMSK modulation scheme; we are simply neglecting the ISI introduced by the *phase*-shaping pulse.

2) The probability density function of the observation x conditioned on state j is

$$b_j(x) = p(x|\mathbf{s}_j) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{|x-m_j|^2}{2\,\sigma^2}\right), \qquad m_j = f(\mathbf{s}_j), \quad 1 \le j \le N \tag{5}$$

$$\mathbf{B} = [b_1(x), \cdots, b_N(x)]^T. \tag{6}$$

3) The state transition probability distribution is

$$\mathbf{A} = \{a_{ij}\}, \qquad 1 \le i, j \le N$$
$$a_{ij} = P(\mathbf{s}[n+1] = \mathbf{s}_j \,|\mathbf{s}[n] = \mathbf{s}_i) = \begin{cases} P(s_j^{(1)}), & \text{if } s_j^{(k)} = s_i^{(k-1)}, \quad k = 2, \cdots, L \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $s_j^{(k)} \in \{-1, 1\}$ denotes the kth element (symbol) in $\mathbf{s}_j = [s_j^{(1)}, s_j^{(2)}, \cdots, s_j^{(L)}, \theta_j]^T$.

The initial state distribution vector $\pi = [\pi_1, \cdots, \pi_N]^T$, where $\pi_i = P(\mathbf{s}[0] = \mathbf{s}_i)$ is assigned an arbitrary value, say, $\pi_i = 1/N$. In the sequel, we will assume that L is either known or can be upper bounded and that symbols are equally likely. Then, we can follow standard HMM-based approaches to estimate the unknown parameter of the model, namely, $\mathbf{m} = [m_1, m_2, \cdots, m_N]^T$—the ISI-corrupted received signal corresponding to each state of the transmitter—and σ^2.

A. The BW Identification Algorithm

Maximum-likelihood estimation using the Baum–Welch (BW) algorithm is the most common solution to the problem of blindly estimating the unknown parameters of the HMM. This iterative method is known to lead, at least, to a local maximum of the likelihood function (e.g., [11]).

Let us consider a block of D samples of the received sequence $\mathbf{x}_D = (x[1], x[2], \cdots, x[D])^T$. First of all, we will obtain $\gamma_i[n]$ $(i = 1 \cdots N;\ n = 1 \cdots D)$, i.e., the probability of being in state $\mathbf{s}_i$ at time n, given $\mathbf{x}_D$ and the model, by means of the computationally efficient *forward–backward* algorithm [11]. Second, we will reestimate the parameters of the model using the BW reestimation formulas. For each state, the mean is estimated weighting every observation with the probability of being in such a state and averaging along the D observations. The estimate for the variance of the noise is derived in a similar manner:

$$\hat{m}_i = \frac{\sum_{n=1}^{D} \gamma_i[n] x[n]}{\sum_{n=1}^{D} \gamma_i[n]}, \qquad 1 \le i \le N \tag{8}$$

$$\hat{\sigma}^2 = \frac{1}{D} \sum_{n=1}^{D} \sum_{i=1}^{N} \gamma_i[n] |\hat{m}_i - x[n]|^2. \tag{9}$$

The above procedure is repeated until a stable point is attained. Once the BW iteration is over, data detection can be performed following: 1) an individually most likely state criterion (IMLS) or 2) a most likely state sequence criterion. There might be some problems with the former in those cases that there are disallowed transitions (i.e., $a_{ij} = 0$ for some i, j) because the obtained state sequence might be impossible. In practice, the problem outlined for criterion 1) does not usually occur [11] so that, for the sake of simplicity, we chose the first criterion to carry out sequence detection. Note, however, that data detection is a side process in the estimation loop. Detected bits are not used in the reestimation formulas but $\gamma_i[n]$ and, hence, this simplification does not degrade convergence properties of the algorithm.

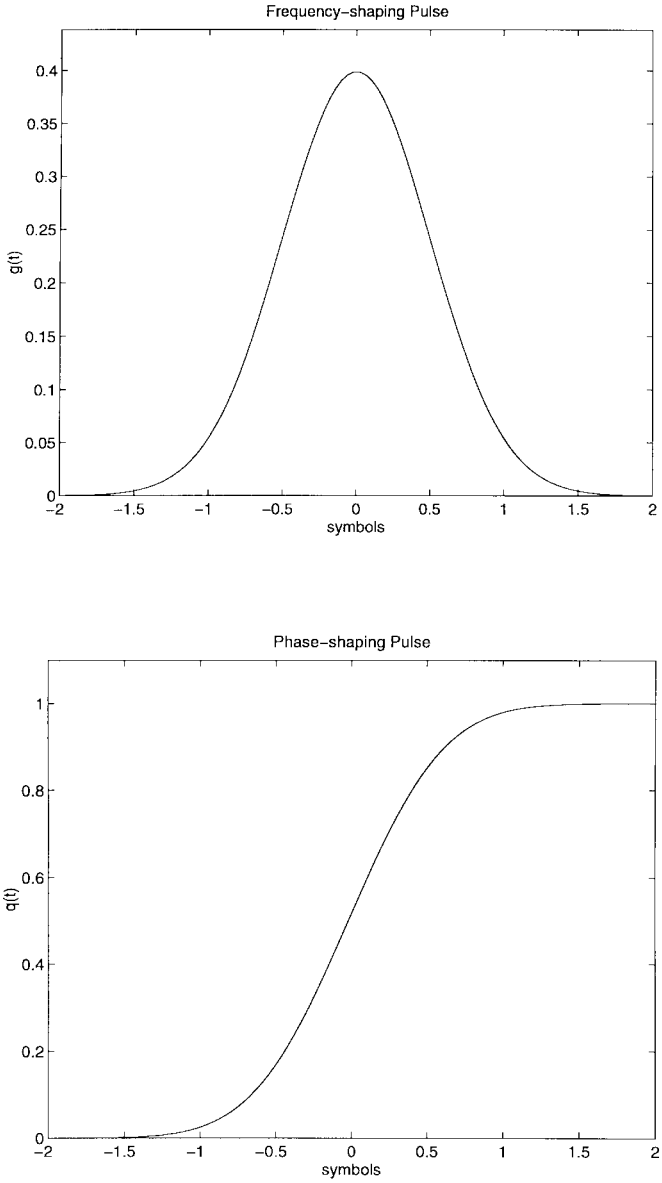

Fig. 3. Frequency and phase shaping filter responses.

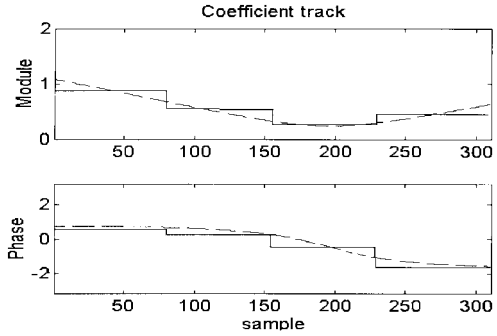

Fig. 4. Tracking for the first tap of the CIR versus time in amplitude and phase. SBBW algorithm (test channel: RA250).

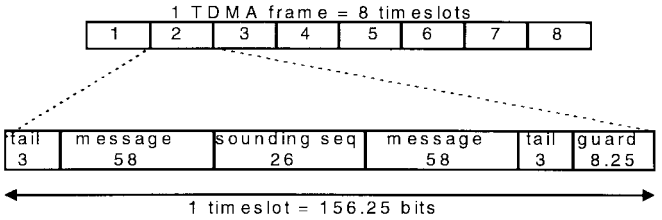

Fig. 5. TDMA frame and normal burst structure in the GSM system.

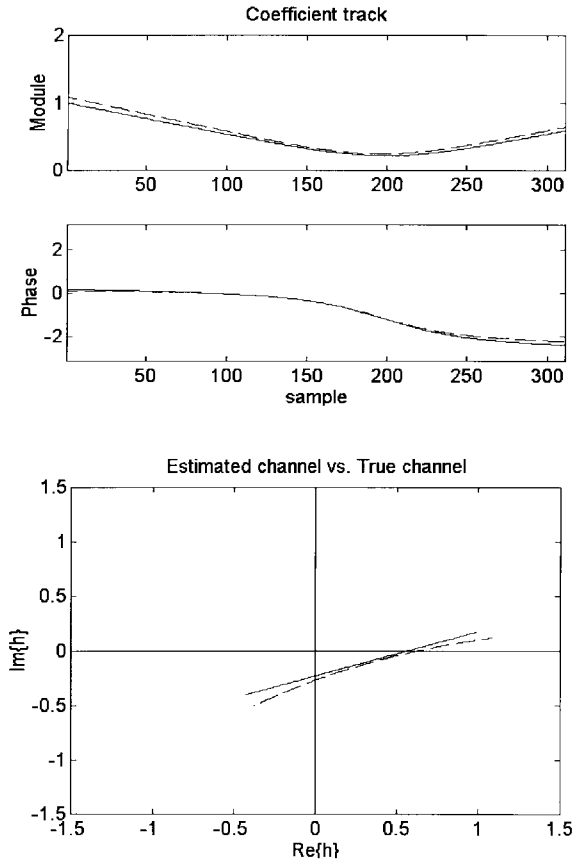

Fig. 6. Tracking for the first tap of a RA250 chanel versus time in amplitude and phase (top) and in rectangular coordinates $\Re\{h_1(t)\}$ and $\Im\{h_1(t)\}$ (bottom). Dashed lines stand for the true channel; solid lines stand for the TDBW estimate.

B. Linear Channels

We observe that the BW algorithm provides a ML approach to the blind identification problem in the general nonlinear case. If, as in our case, a parametric model of the channel is assumed and the modulation scheme is also known, we can use this information to improve parameter estimation with a constrained optimization. For example, for a linear (FIR) channel, we can include the following linear constraint:

$$\mathbf{m} = \mathbf{D}\,\mathbf{h} \tag{10}$$

where $\mathbf{h} = [h_0, h_1, \cdots, h_{L_c-1}]^T$ stands for the channel impulse response (CIR), and $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_N]^T$ is a $N \times L_c$ matrix containing in its rows all the L_c-tuples $\mathbf{d}_i = [d_i^{(1)}, d_i^{(2)}, \cdots, d_i^{(Lc)}]^T$ corresponding to the modulator consecutive outputs associated with the N different states of the system. After each reestimation of

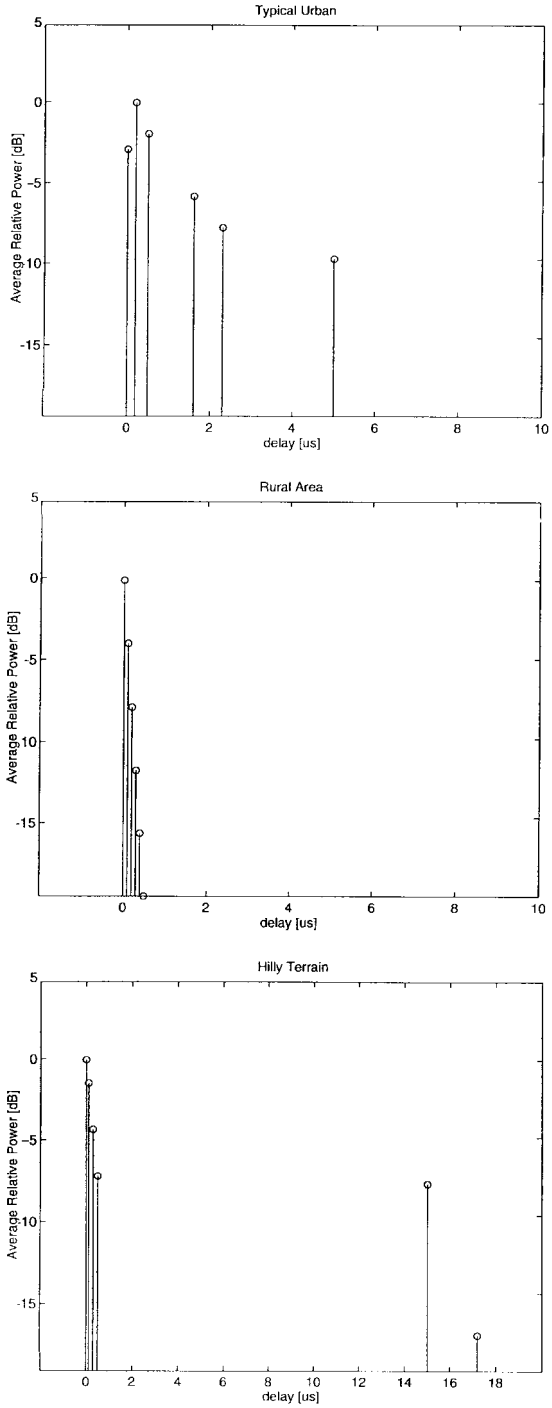

Fig. 7. Six-coefficient propagation models standardized by the ETSI.

$\mathbf{m} = [m_1, m_2, \cdots, m_N]^T$, they are projected with

$$\hat{\mathbf{m}} = \mathbf{D}\,\mathbf{D}^{\#}\hat{\mathbf{m}}$$

or

$$\hat{\mathbf{m}} = \mathbf{D}\,\hat{\mathbf{h}} \tag{11}$$

where $\mathbf{D}^{\#}$ denotes the pseudoinverse of $\mathbf{D}$, and $\hat{\mathbf{h}} = \mathbf{D}^{\#}\hat{\mathbf{m}}$ is a *least squares* (LS) estimate of $\mathbf{h}$ from $\hat{\mathbf{m}}$. It is worth mentioning that this estimate for $\mathbf{h}$ assumes that the estimation errors observed in the components of $\mathbf{m}$ are independent and with equal variance. For CIR's exhibiting large delay spread, the number of states in the model (N) increases rapidly, whilst the number of states observed in

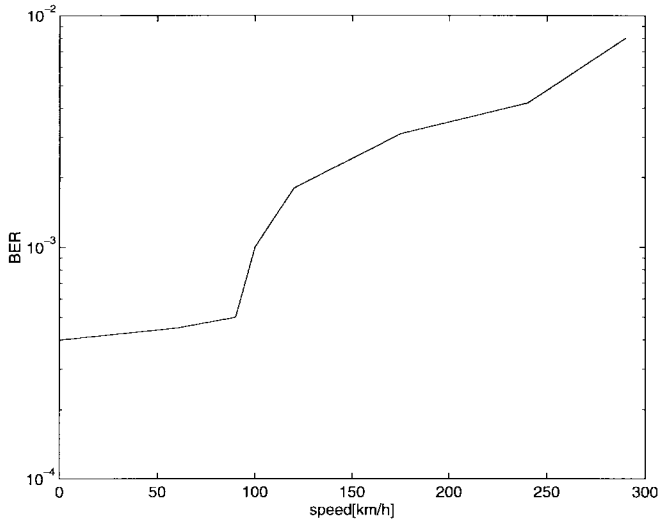

Fig. 8. Performance of BBW algorithm versus the speed of the mobile (TU-type chanel with $E_b/N_o = 24$ dB).

a timeslot period (D) remains constant. As a result, the variance in the estimation of some components in the means vector increases, and this fact leads to a severe distortion in the CIR estimate. Consequently, we replace the LS estimate of $\mathbf{h}$ by a weighted least squares (WLS) estimate considering the following (diagonal) weighting matrix:

$$\mathbf{W} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & w_N \end{pmatrix}$$

$$w_i = \sum_{n=1}^{D} \gamma_i[n]. \tag{12}$$

In other words, if the estimate for component m_i is not reliable because that state was seldom observed (and, hence, w_i is small), the error committed in that component is not considered in future reestimates. From now on, this algorithm will be referred as the batch-BW (BBW) algorithm.

Regarding identifiability issues, it is well known that hte BW algorithm suffers from local maxima. However, we have observed that the inclusion of the linear constraints in (10) helps in fighting against such a problem. Time-shift ambiguity is observed as in any blind procedure where no temporal reference is given to the receiver. On the contrary, scale ambiguity is not observed except for a phase shift inherent to any blind scheme applied to symmetric constellation and/or symbol alphabets.

III. Techniques for Varying Channels

It is clear that the BBW algorithm presented in the preceding section implicitly assumes the CIR to be stationary within a timeslot. A first approach to the solution consists in splitting up timeslots in several subblocks producing different CIR estimates in each (Fig. 4). The CIR estimated in a subblock will be used as the initial estimate for the following one. The resulting algorithm will be referred as the segmented batch-BW (SBBW) algorithm. Nevertheless, there is a problem in segmenting data in subblocks: The CIR must be estimated with less data. This can be compensated with the inclusion of the weighting matrix but, in any case, overfragmentation should be avoided because it reverts in a higher BER.

Recursive methods can be developed, but as long as a burst-oriented TDMA access is considered in the GSM system (Fig. 5), a burst-oriented detection is more suitable. Additionally, at the chosen bit rate (270.8 kb/s), multipath propagation as well as Doppler lead to deep fades, which might lead adaptive algorithms to lose channel tracking [13].

A. Including Time Dependence in the HMM

For the above reasons, we include the time-varying nature of the channel directly in the batch reestimation formulas. In fact, we can approximate the evolution of every tap in the CIR estimate

$$\hat{\mathbf{h}}[n] = \left[\hat{h}_{(0)}[n] \cdots \hat{h}_{L_c-1}[n]\right]^T$$

with a polynomial in n:

$$\hat{\mathbf{h}}[n] = \mathbf{h}^{(0} + \mathbf{h}^{(1} \cdot n + \mathbf{h}^{(2} \cdot n^2 + \mathbf{h}^{(3} \cdot n^3 \cdots \tag{13}$$

For the channels specified in the ETSI recommendation and the range of velocities considered for the mobile station (less than 250 km/h), the first-order approximation was observed to be accurate enough. Applying the linear transform in (10), we get

$$\begin{aligned} \hat{\mathbf{m}}[n] &= \mathbf{D}\,\hat{\mathbf{h}}[n] \\ &= \mathbf{D}\,(\mathbf{h}^{(0} + \mathbf{h}^{(1} \cdot n) \\ &= \mathbf{m}^{(0} + \mathbf{m}^{(1} \cdot n. \end{aligned} \tag{14}$$

Vectors $\mathbf{m}^{(0}$ and $\mathbf{m}^{(1}$ will be derived as the ones that minimize the MSE given by the following expression:

$$e_i^2 = \frac{\sum_{n=1}^{D} \gamma_i[n]\,|x[n] - m_i^{(0} - m_i^{(1} \cdot n|^2}{\sum_{n=1}^{D} \gamma_i[n]}, \quad 1 \le i \le N \tag{15}$$

$$\varepsilon^2 = \sum_{i=1}^{N} e_i^2. \tag{16}$$

Differentiating each component in the sum ε^2 with respect to $m_i^{(0}$ and $m_i^{(1}$, we obtain

$$\begin{aligned} \nabla e_i^2 &= \begin{pmatrix} \dfrac{\partial e_i^2}{\partial m_i^{(0}} \\ \dfrac{\partial e_i^2}{\partial m_i^{(1}} \end{pmatrix} \\ &= -2 \cdot \begin{pmatrix} \Re\left\{\sum_{n=1}^{D} \gamma_i[n](x[n] - m_i^{(0} - m_i^{(1} \cdot n)\right\} \\ \Re\left\{\sum_{n=1}^{D} \gamma_i[n](x[n] - m_i^{(0} - m_i^{(1} \cdot n) \cdot n\right\} \end{pmatrix} \end{aligned} \tag{17}$$

whose Hessian

$$\begin{aligned} H_{e_i} &= 2 \cdot \begin{pmatrix} \sum_{n=1}^{D} \gamma_i[n] & \sum_{n=1}^{D} \gamma_i[n] \cdot n \\ \sum_{n=1}^{D} \gamma_i[n] \cdot n & \sum_{n=1}^{D} \gamma_i[n] \cdot n^2 \end{pmatrix} \\ &\doteq \begin{pmatrix} A & B \\ B & C \end{pmatrix} \end{aligned} \tag{18}$$

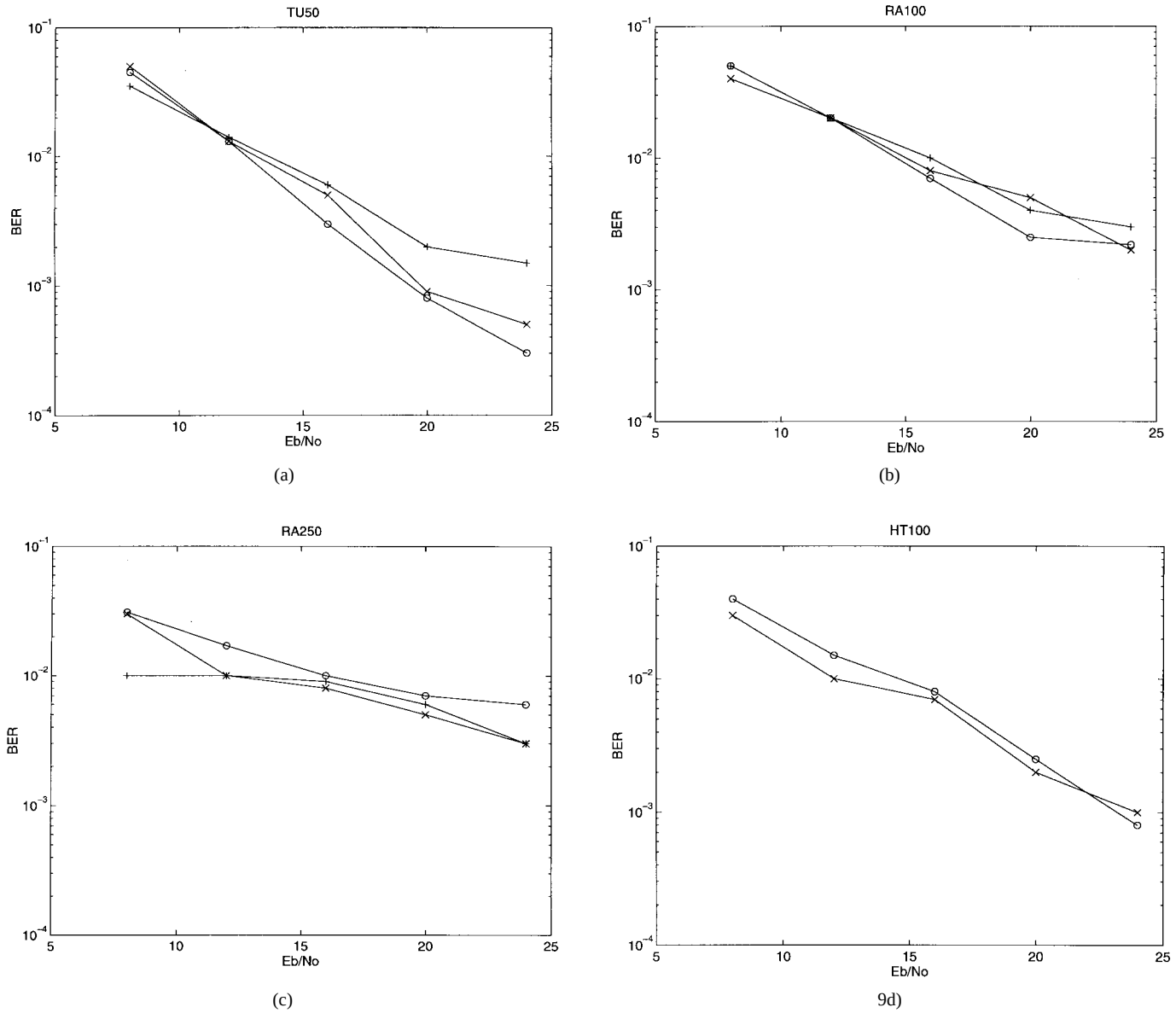

Fig. 9. Comparative performance for different test channels: (o) Viterbi-based nonblind receiver, (x) SBBW, (+) TBDW.

is positive definite since the determinants

$$\begin{aligned}\Delta_1 &= A \\ &= \sum_{n=1}^{D} \gamma_i[n] \\ \Delta_2 &= AC - B^2 \\ &= \sum_{n=1}^{D} \gamma_i[n] \cdot \sum_{n=1}^{D} \gamma_i[n] \cdot n^2 - \left(\sum_{n=1}^{D} \gamma_i[n] \cdot n \right)^2 \end{aligned} \tag{19}$$

are both strictly positive unless c1) that state was not observed or c2) was observed only once for $n = n_i$:

$$\text{c1)} \quad \sum_{n=1}^{D} \gamma_i[n] = 0 \tag{20}$$

$$\text{c2)} \quad \sum_{n=1}^{D} \gamma_i[n] = \gamma_i[n_i] \tag{21}$$

and, consequently, there is no sense in looking for a linear approximation. Equating the gradient to zero and carrying out proper transformations, we find that

$$m_i^{(0} = \frac{A \cdot \left(\sum_{n=1}^{D} \gamma_i[n] x[n] \cdot n \right) - B \cdot \left(\sum_{n=1}^{D} \gamma_i[n] x[n] \right)}{\Delta_2}, \quad 1 \le i \le N \tag{22}$$

$$m_i^{(1} = \frac{C \cdot \left(\sum_{n=1}^{D} \gamma_i[n] x[n] \cdot n \right) - B \cdot \left(\sum_{n=1}^{D} \gamma_i[n] x[n] \cdot n \right)}{\Delta_2}, \quad 1 \le i \le N \tag{23}$$

$$\hat{\sigma}^2 = \frac{1}{D} \sum_{n=1}^{D} \sum_{i=1}^{N} \gamma_i[n] | x[n] - m_i^{(0} - m_i^{(1} \cdot n |^2 \tag{24}$$

provide the components of the desired vector and an estimate for the variance of the AWGN.

Finally, special measures should be taken for the cases above mentioned. In the first case c1), we block the contribution of those components in the means vector with a weighting matrix as described in Section II. In the second case, the *static* estimate for $\mathbf{m}$ replaces the linear approximation. In the sequel, this version will be referred as the time dependent BW (TDBW) algorithm.

Channel tracking properties for TDBW version are shown in Fig. 6. The TDBW version is far more robust against deep fades than other adaptive versions. The reason for this robustness being that TDBW is a batch algorithm, where every sample in the timeslot is used to estimate the CIR in every instant (even though in deep local fades), whereas in the adaptive versions, the estimate relies mainly on the previous and, maybe, already-faded samples. The main drawback of the TDBW version is the increase in the computational burden when compared with the BBW version. Moreover, in the case of TDBW, the number of parameters to be estimated doubles the quantity required before since vectors $\mathbf{m}^{(0}$ and $\mathbf{m}^{(1}$, instead of only vector $\mathbf{m}$, must be estimated now. As before, the dimension of those vectors depends on the number of states in the model depending and, in turn, on the CIR delay spread. Therefore, longer timeslots might be required for CIR's exhibiting large delay spreads in order to allow the algorithm to converge.

IV. Results

Algorithms were tested with the six-coefficient propagation models defined by the European Telecommunications Standard Institute (ETSI), which correspond to typical scenarios [13]: rural area (RA) channel—the least hostile in terms of ISI—typical urban (TU), and hilly terrain (HT) channels (average power profiles are plotted in Fig. 7). Besides, Doppler frequency shifts due to the speed of the mobile station (MS) are also considered in these channel models: a given test channel, for example, TU50, is defined not only by its channel type—TU—but also by the speed of the mobile − 50 km/h. The speed of the MS for each environment was chosen according to [14]. With respect to the ability to overcome ISI, the most interesting channel was HTx since it exhibits the longest delay spread. Concerning the tracking properties of the algorithms, the most interesting cases were RA250 and RA100 because channel coherence time intervals were the lowest ones. A sampling rate of two samples per symbol was employed in the simulations in order to avoid using matched filtering [15]. For evaluating bit error rates, 200-timeslot simulations (31 200 symbols) were performed.

A. Comparative Performance

In Fig. 8, we plot, for the BBW version, the BER against the speed of the mobile for a TU-type channel. For speeds above 80–100 km/h, a deep degradation in performance is observed. Hence, beyond this threshold, it would be necessary to consider one of the developed versions mentioned before.

In Fig. 9, performance for SBBW and TDBW versions is compared by means of computer simulations with that exhibited by the Viterbi-based receiver proposed in [13] and [14] for different E_b/N_o ratios. This Viterbi-based receiver does not have knowledge of the CIR, but it *estimates* such a CIR using the training sequence embedded in the burst. Therefore, error rates achieved with such algorithm should not be considered to be absolute benchmarks. Performance for our set of algorithms is very similar to that exhibited by the Viterbi-based receiver. However, the comparison is not straightforward in the case of RA channels. The reason is that in the original paper [14], Rayleigh statistics were assumed for all coefficients, whereas in our study and according to the ETSI standard, a Rice pdf is considered for the first one. It is equivalent to admitting a direct line of sight, which is far more realistic in such scenarios.

Performance loss due to estimating twice the number of parameters when considering the TDBW scheme is clearly observed in the TU50 case. Moreover, in the HT test channel, the TDBW algorithm fails to converge as a consequence of the lack of sufficient data. Finally, note that the TDBW version is revealed to be more useful as long as the speed of the mobile increases (i.e., RA250). In that case, it clearly outperforms the SBBW version.

V. Conclusions

In this paper, we have applied the theory of hidden Markov models, which have been thoroughly studied in speech recognition applications, to the communications problem of blind sequence detection and channel estimation. A HMM has been built for the GSM system and, on the basis of the BBW algorithm, two algorithms (SBBW and TDBW) have been developed in order to fight against the impairments observed in the radio channel.

Performance has been evaluated for the GSM system and has been found to be close to that achieved by nonblind schemes. Moreover, a blind receiver based on such algorithms would require no training sequences, and that would imply a $26/156 = 17\%$ increase in the capacity of voice bursts (see Fig. 5). On the other hand, the most important drawback of the algorithms is their high computational burden, although the nonblind reference receiver [14] is rather sophisticated as well. Three trends can be outlined for future work. First, we have a detailed analysis of the computational complexity and convergence rate for the developed algorithms. Second, we have the inclusion of spatial diversity in our model to fight against deep fades. Finally, applying the developed BW-based algorithms in other communication environments such as CDMA.

References

[1] J. G. Proakis and C. L. Nikias, "Blind equalization," in *Proc. SPIE Adaptive Signal Processing*, vol. 1565, 1991, pp. 76–87.

[2] J. Shynk *et al.*, "A comparative performance study of several blind equalization algorithm," in *Proc. SPIE Adaptive Signal Processing*, vol. 1565, 1991, pp. 102–117.

[3] J. M. Mendel, "Tutorial on higher-order statistics (Spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 277–305, Mar. 1991.

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

[5] J. A. R. Fonollosa and J. Vidal, "Application of hidden Markov models to blind channel characterization and data detection," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Procesisng*, Australia, Apr. 1994, pp. 185–188.

[6] J. R. Fonollosa *et al.*, "Blind multiuser detection with array observation," *Wireless Personal Commun.-Special Issue on Interference Mobile Wireless Syst.*, 1996, to appear.

[7] M. Erkurt and J. G. Proakis, "Joint data detection and channel estimation for rapidly fading channels," in *Proc. IEEE Global Telecommun. Conf. (Globecom '92)*, 1992, pp. 910–914.

[8] M. Feder, and J. A. Catipovic, "Algorithms for joint channel estimation and data recovery—Application to equalization in underwater communications," *IEEE J. Oceanic Eng.*, vol. 16, no. 1, pp. 42–55, 1991.

[9] R. Vallet, "Symbol by symbol MAP detection and the Baum–Welch identification algorithm in digital communications," in *Proc. Europ. Signal Processing Conf.*, Sept. 1992, pp. 131–134.

[10] M. Seshadri, "Joint data and channel estimation using fast blind trellis search techniques," *IEEE Trans. Commun.*, vol. COM-42, pp. 1000–1011, Feb./Mar./Apr. 1994.

[11] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[12] H. Murota and K. Hirade, "GMSK modulation for mobile radio telephony," *IEEE Trans. Commun.*, vol. COM-29, no. 7, pp. 1044–1050, July 1981.

[13] R. Steele, *Mobile Radio Communications.* London, U.K.: Pentech, 1992, pp. 677–768.
[14] R. d'Avella, L. Moreno, and M. Sant'Agostino, "An adaptive MLSE receiver for TDMA digital mobile radio," *IEEE J. Selected Areas Commun.*, vol. 7, no. 1, pp. 122–129, Jan. 1989.
[15] G. Ungerboeck, "Fractional tap spacing equalizer and consequences for clock recovery in data modems," *IEEE Trans. Commun.*, vol. COM-24, pp. 856–864, Aug. 1976.

A Blind Equalizer for Nonstationary Discrete-Valued Signals

Ta-Hsin Li and Kais Mbarek

***Abstract*— Adaptive algorithms are proposed for blind equalization of communication channels. The algorithms explicitly utilize the finite alphabetical set of the input signals and minimize a criterion that depends solely on the alphabetical set. The method is shown to be able to handle nonstationary signals without requiring or estimating their time-varying statistical parameters. Simulation results are presented to test and demonstrate the method.**

I. INTRODUCTION

In digital communications, blind equalizers are employed to correct linear distortions caused by the transmission channel while training signals are not available [1]. Recently proposed adaptive algorithms for on-line adjustment of blind equalizers are either based on the inverse filtering approach (e.g., [2]–[4]) or derived from the method of moments (e.g., [5], [6]). These blind equalizers are general procedures that can be used to recover a large class of non-Gaussian signals. But when applied to digital communication systems, they all ignore an important fact: The transmitted signals take values only in a discrete and finite alphabetical set that is known at the receiver. This important *a priori* information has been proven extremely useful. For instance, using this information, a Bayesian approach can be easily employed to handle noisy data and nonstationary correlated signals with unknown statistical parameters [7].

Within the non-Bayesian inverse filtering framework, the discreteness information can lead to improved accuracy for channel estimation. The block-processing method in [8]–[10] is such an example. To estimate the inverse system impulse response, this method seeks a linear filter to minimize a criterion that measures the closeness of the filter output to a discrete-valued signal with the known alphabets. For the ARMA channels, the method achieves a mean-squared error that goes to zero as quickly as an exponential function of the sample size; whereas, without using the discreteness, the error usually decays only as a function of the reciprocal of the sample size. A faster error decaying implies more accurate channel estimation and signal recovery. Another advantage of the method is its ability to handling nonstationary signals without using or estimating their statistical parameters. This property simplifies the implementation and is especially preferable in situations where the statistical parameters of the transmitted signals vary rapidly over time.

Manuscript received December 1, 1995; revised August 28, 1996.

The authors are with the Department of Statistics, Texas A&M University, College Station, TX 77843-3143 USA (e-mail: thl@stat.tamu.edu).

Publisher Item Identifier S 1053-587X(97)00645-4.

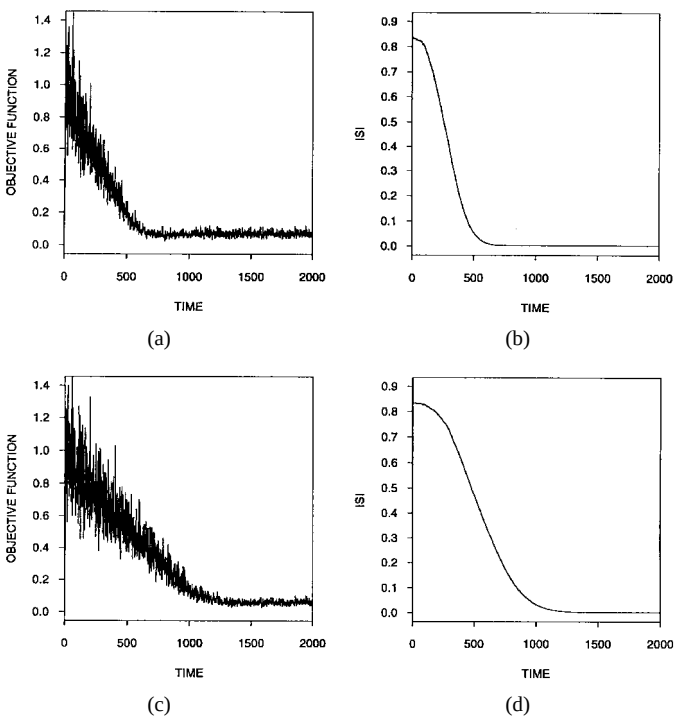

Fig. 1. Telephone channel with the stationary equiprobable PAM(2) signal in Example 1. (a)–(b) Average trajectory of the objective function and of the ISI for $\mu = 2 \times 10^{-3}$. (c)–(d) Same average trajectories for $\mu = 10^{-3}$. All trajectories are obtained from 30 independent runs of (4) with $M = 6$.

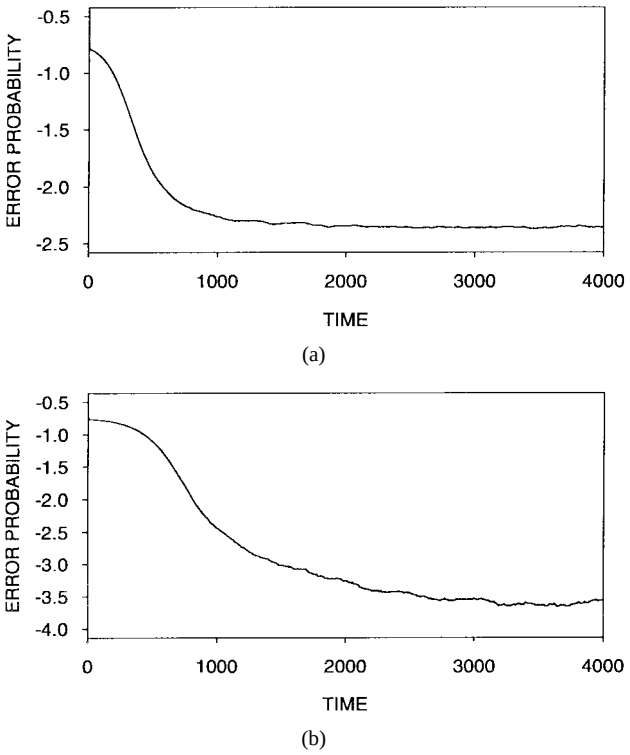

Fig. 2. Probability of error for the telephone channel with the stationary equiprobable PAM(2) signal in Example 1. The trajectories, shown in logarithmic scale of base 10, are based on 1000 independent runs of (4) with $M = 6$ and are smoothed by moving average of length 200. (a) $\mu = 2 \times 10^{-3}$. (b) $\mu = 10^{-3}$.