APPLICATION OF HIDDEN MARKOV MODELS TO BLIND CHANNEL ESTIMATION AND DATA DETECTION IN A GSM ENVIRONMENT

Carles Antón-Haro, José A. R. Fonollosa and Javier R. Fonollosa.
Dpt. of Signal Theory and Communications. Universitat Politècnica de Catalunya.
c/ Gran Capità s/n. 08034 Barcelona (SPAIN)
Tel: +34-3-4016454, Fax: +34-3-4016447, e-mail: carles@gps.tsc.upc.es

ABSTRACT

In this paper, we present an algorithm based on the Hidden Markov Models (HMM) theory to solve the problem of blind channel estimation and sequence detection in mobile digital communications. The environment in which the algorithm is tested is the Paneuropean Mobile Radio System, also known as GSM. In this system, a large part in each burst is devoted to allocate a training sequence used to obtain a channel estimate. The algorithm presented would not require this sequence, and that would imply an increase of the system capacity. Performance, evaluated for standard test channels, is close to that of non-blind algorithms.

1. INTRODUCTION

It is well known that no high-speed band-limited digital communication can be carried out without the help of an equalizer. Conventional approaches to the adjustment of this equalizer require the transmission of a *training sequence* (i.e. known *a priori* by the receiver and the transmitter), which provides an accurate initial estimate for the equalizer taps; afterwards, slighter adjustments can be made on a decision directed (DD) basis to adapt this first estimate to the, almost always, changing environment. Of course, the transmission of these training sequences, when possible, brings down the capacity of the system. For that reason, there is an increasing interest around blind equalizers [1,2,3] which deal with the problem of the adjustment *without* training sequences (i.e. *blindly*).

In [3], an Estimation-Modification (EM) Viterbi-based algorithm is proposed to perform jointly a Maximum Likelihood (ML) channel estimation and sequence detection. However, modeling the received signal as a HMM allows us to make use of the complete theory developed for these models. For example, the Baum&Welch (BW) algorithm was proposed in [7] to estimate the parameters of the channel and the characteristics of the modulation. This algorithm is known to lead, at least, to a local maximum of the likelihood function [4], what is not guaranteed by the Viterbi algorithm. In this paper, several modifications to this previously proposed algorithm are introduced to deal with mobile radio channels.

This work was supported by the Comissionat per a les Universitats i Recerca (Generalitat de Catalunya).

2. OVERVIEW OF THE GSM SYSTEM

The whole GSM system is specified in the ETSI (European Telecommunication Standards Institute) recommendations. We are not concerned in a full study of it, but in those points related to the transmission subsystem.

2.1 Transmission Subsystem

The GSM system operates in the 900 MHz band. A constant-envelope Gaussian Minimum Shift Keying (GMSK) modulation scheme with a relative bandwidth (BT) equal to 0.3 is used, due to its good spectral properties. The access strategy is TDMA with 8 timeslots per carrier. At the chosen bit rate (270.8 kb/s), multipath propagation leads to deep fades and to uncontrolled Intersymbol Interference (ISI). Besides, receiver's mobile nature allows Doppler effect to show up. As depicted in Fig. 1, each timeslot in Normal Bursts reserves its 26 central bits to host a sounding sequence. This sequence is used at the receiver to obtain an estimate of the channel impulse response (CIR). This information is then used by the detector (i.e. a MLSE-type) to perform data detection. In contrast, a blind equalizer would make this sequence unnecessary, so that more voice channels could be allocated instead.

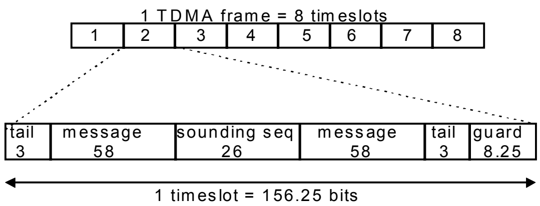

Fig. 1: TDMA frame and Normal Burst structure in GSM system

2.2 Channel Models

In order to provide identical test conditions for different implementations of the GSM system, a set of 12-echo and 6-echo propagation models, corresponding to typical scenarios were defined by the ETSI (see Fig. 2) [5]. Delay spreads for different signal reflections are comprised among 0.5μs in the Rural Area (RA) channel - the least hostile in terms of ISI -, and 17.5μs for the Hilly Terrain (HT) one. Taking into account the bit interval duration (3.60 μs), we observe that channel memory is up to 4 symbols. For reliable communication, the uncontrolled ISI introduced by the channel as well as that introduced deliberately by the partial response modulator [6], have to be removed by the channel equalizer.

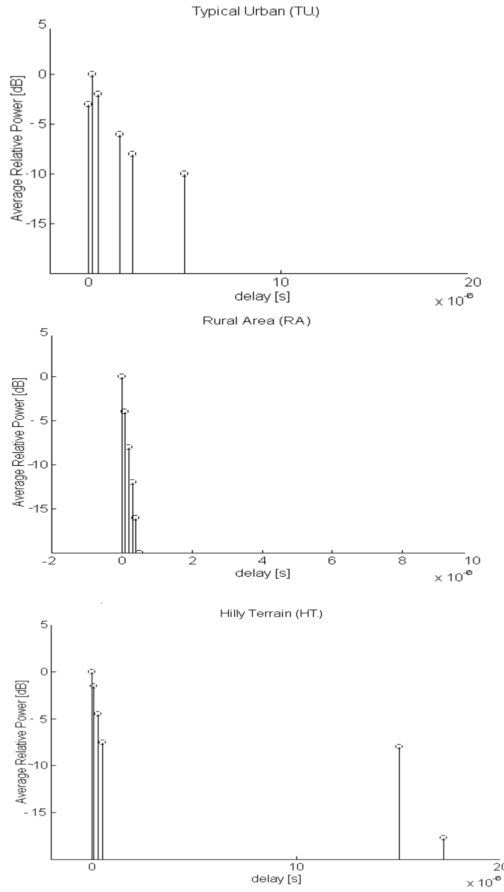

Fig. 2: Some of the 6-echo propagation models defined in the ETSI recommendation

3. SIGNAL MODEL

According to section 2, the signal at the input of the BW detector can be modeled as:

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

where f(.) is a nonlinear function of the present state $\mathbf{s}[n]$, and w[n] denotes a sequence of zero-mean Gaussian variables with variance σ^2.

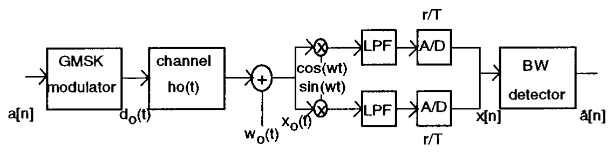

Fig. 3: Transmission subsystem.

Note that: (1) the GMSK modulator is nonlinear, and (2) both the modulator and the channel contribute to the system memory. Assuming a linear model for the channel, f can be expressed as:

$$\mathrm{f}(\mathbf{s}[n]) = \sum_{i=0}^{L_c-1} h[i]\, d[n-i] = \sum_{i=0}^{L_c-1} h[i]\, \mathrm{e}^{\mathrm{j}\cdot\phi[n-i]} \tag{2}$$

where h and d are the baseband equivalences for h_0 and d_0. For a modulation index equal to 0.5, $\phi[n]$ states as:

$$\phi[n] = \pi \sum_{r=-R}^{R} q[r]\, a[n-r] + \theta[n] \tag{3}$$

where $q[r] \in [0,0.5]$ are the weights corresponding to the (sampled) gaussian-shaping pulse, and $\theta[n] \in \{0,\pi/2,\pi,3\pi/2\}$ accounts for the accumulated phase at instant n. Now we conclude that the number of transmitter symbols involved in a single observation at the receiver is given by:

$$L = L_m + L_c - 1 = (2R+1) + L_c - 1 \tag{4}$$

However, the amount of ISI produced by the GMSK modulator for BT=0.3 can be neglected without significant performance loss. Under this simplifying assumption, that reduces significantly the number of states and computational complexity, we get that $R=0 \Rightarrow L=L_c$. At this point, we can model the observations as a probabilistic function of the state $\mathbf{s}[n]=(\mathrm{a}[n],..,\mathrm{a}[n-L+1],\theta[n])^{\mathrm{T}}$, obtaining a description of $\mathbf{x}_D=(x[1],x[2],..,x[D])^{\mathrm{T}}$ as a first order HMM with the following characteristics:

1. Nr of states: $N=4\cdot 2^L$. We denote individual states as $\mathbf{S}=(\mathbf{s}_1,...,\mathbf{s}_N)^{\mathrm{T}}$ and $\mathbf{s}$[n] as the state at instant n.
2. The probability density function conditioned to state j is:

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

3. The state transition probability distribution is:

$$\mathbf{A} = \{a_{ij}\}, \quad 1 \le i,j \le N$$

$$a_{ij} = P(\mathbf{s}[n+1] = \mathbf{s}_j \,|\, \mathbf{s}[n] = \mathbf{s}_i) = \tag{6}$$
$$= \begin{cases} P(s_j^{(1)}) & , \ if\ s_j^{(k)} = s_i^{(k-1)},\ k = 2..L \\ 0 & ,otherwise \end{cases}$$

where $s_j^{(k)} \in \{-1,1\}$ denotes the k^{th} element in $\mathbf{s}_j$:

$$\mathbf{s}_j = (s_j^{(1)}, s_j^{(2)}, .., s_j^{(L)}, \theta)^T \tag{7}$$

4. THE BW ALGORITHM

Once the HMM is built, the BW algorithm gives us a solution to the problem of identificating the unknown parameters of the model: $\mathbf{m}=(m_1...m_N)^T$ (or equivalently $\mathbf{h}=(h_0...h_{Lc-1})^T$), and σ^2. A solution to the problem of sequence detection is also obtained [4,7]. It is worth making a point of the additional constraints than can be added as a result of the assumption of a FIR model for the channel. In fact, after each reestimation for the parameters of the model, it is possible to project them in the following manner [7] :

$$\hat{\mathbf{m}} = \mathbf{D}\,\mathbf{D}^{\#}\,\hat{\mathbf{m}} \qquad (8)$$

where $\mathbf{D}=(\mathbf{d}_1,\mathbf{d}_2,..,\ \mathbf{d}_N)^T$ is a NxL_c matrix containing in its rows all the possible L_c-tuples of the modulator consecutive outputs, $\mathbf{d}_i=(d_i^{(1)}...d_i^{(Lc)})^T$, corr esponding to the N different states of the system. $\mathbf{D}^{\#}$ denotes pseudoinverse. Otherwise, in two steps:

$$\hat{\mathbf{h}} = \mathbf{D}^{\#}\hat{\mathbf{m}}$$
$$\hat{\mathbf{m}} = \mathbf{D}\,\hat{\mathbf{h}} \qquad (9)$$

That is to say, from $\mathbf{m}$ we can obtain a Least Squares (LS) estimate for the CIR in that timeslot.

In conclusion, the whole algorithm states as follows:

1. Estimation of $\mathbf{m}$ by means of the additional linear constraint:

$$\hat{\mathbf{m}} = \mathbf{D}\,\hat{\mathbf{h}} \qquad (10)$$

1. BW algorithm itself (see [7] for further details):

   - *Compute the variable* $\gamma_i\,[n]$ (i=1..N, n=1..D), that is, the probability of being in state i at instant n given the whole sequence $\mathbf{x}_D=(x[1],..,x[D])^T$ and the model.
   - *Reestimation of the parameter set* by means of time averaging:

$$\hat{m}_i = \frac{\sum_{n=1}^{L}\gamma_i[n]\,x[n]}{\sum_{n=1}^{L}\gamma_i[n]}, \qquad 1 \le i \le N \qquad (11)$$

$$\hat{\sigma}^2 = \frac{1}{L}\sum_{n=1}^{L}\sum_{i=1}^{N}\gamma_i[n]\left|\hat{m}_i - x[n]\right|^2 \qquad (12)$$

3. Estimation of $\mathbf{h}$ using again linear constraints.

$$\hat{\mathbf{h}} = \mathbf{D}^{\#}\hat{\mathbf{m}} \qquad (13)$$

4. Repeat steps 1..4 until convergence.

5. Sequence detection:

$$\hat{a}[n] = \Psi\left(\underset{1\le i\le N}{\arg max}\left\{\gamma_i[n]\right\}\right) \qquad (14)$$

where $\psi(.)$ is a function that maps each state (i=1..N) onto the present symbol.

5. MODIFIED ALGORITHM

However, when using the above algorithm in a GSM environment, some additional points should be considered:

1. As seen before the original BW is a batch algorithm that obtains a single averaged CIR estimate for every timeslot (D symbols). Therefore, if the CIR varies rapidly a lot of errors will appear.

2. For CIRs exhibiting large delay spreads, the number of states in the model (N) increases rapidly. However the number of states observed in a timeslot period (D), remains constant. As a result, the variance in the estimation of some components in $\mathbf{m}$ increases. That introduces a severe distortion in the CIR estimate and, in some cases, makes the algorithm unstable.

The following strategies were followed, respectively, to cope with these problems:

1. Splitting up timeslots in several subblocks producing different CIR estimates in each. It is also possible to develop recursive versions of the algorithm as the LMS-based one proposed in [7]. Nevertheless, better results were obtained by segmenting timeslots and applying the batch-type BW algorithm in each.

2. Substituting the LS estimation for $\mathbf{h}$ by a Weighted Least Squares (WLS) estimation. That is:

$$\mathbf{D}_{WLS}^{\#} = \left(\mathbf{D}^H\mathbf{W}\,\mathbf{D}\right)^{-1}\mathbf{D}^H\mathbf{W} \qquad (15)$$

   where $^{\#}$ denotes conjugate transpose. The weight matrix was chosen to be:

$$\mathbf{W} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & w_N \end{pmatrix} \qquad (16)$$

   where:

$$w_i = \sum_{n=1}^{L}\gamma_i[n] \qquad (17)$$

   In other words, if the estimate for m_i is not reliable, because that state was seldom observed (w_i small), the error committed in future reestimations of that component is not considered.

With these modifications the algorithm performance was significantly improved.

6. SIMULATION RESULTS

The above algorithm was tested for the channels described in section 2.2. The parameters of the simulation were:

- Sampling rate: 2 samples per symbol to compensate for possible timing errors.
- Initial parameters: $\hat{\sigma}^2 = 1$, $\hat{\mathbf{h}} = [1,0,...,0]^T$
- Subblocks/timeslot: Up to 4, for the RA250 channel.
- Speed of the mobiles: From 50 to 250 km/h.

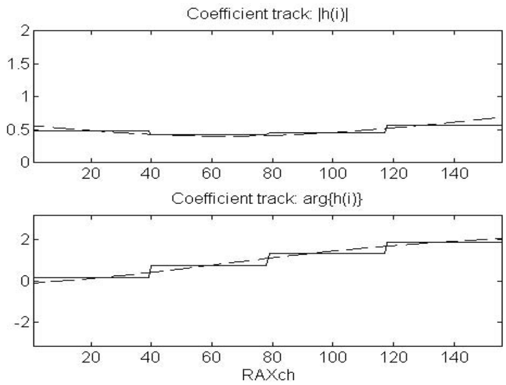

Fig. 4: Tracking for the first tap of the CIR vs time in amplitude and phase when each timeslot is segmented in subblocks (Test channel: RA250).

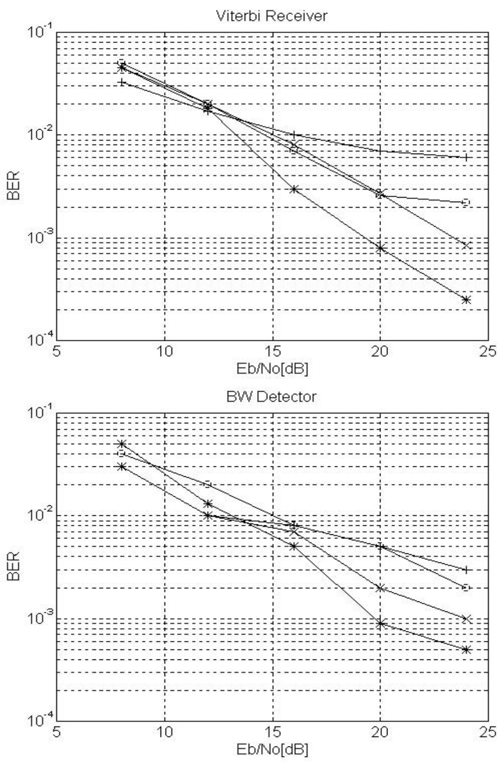

Fig. 5: Comparative performance for different test channels and speed of the mobiles: RA250 (+), RA100 (o), HT100 (x), TU50 (*).

In Fig. 4 tracking properties are shown. As mentioned in section 5, we observe that the channel estimate for each subblock is an average of that really observed. Due to CIR rapid variations, 4 subblocks are required in this case.

Performance compared with that exhibited by the Viterbi-based receiver proposed in [8,9] for different E_b/N_o ratios, is shown in Fig. 5. It is clear how close performance is for both receivers. However, the comparison is not straightforward in the case of RA channels. The reason is that in the reference paper, [8], Rayleigh statistics were assumed for all the echoes, whereas, in our study and according to the ETSI recommendations, a Rice pdf is considered for the first one. It is equivalent to admit a direct line of sight, what is far more realistic in such scenarios.

In the case of HT channel, it is also important to point out that such a satisfactory behaviour is a direct consequence of including the Weighting Matrix.

7. CONCLUSIONS

A BW-based algorithm for blind sequence detection and channel estimation has been presented. Performance, evaluated in a very concrete environment (the GSM system), is close to that achieved by non-blind equalizers. However, the most important drawback of the algorithm is its high computational burden.

Future work is concerned about reducing the computational complexity of the algorithm, as well as developing recursive (and consequently adaptive) versions of it, suitable for this scenario. It is also being studied the possibility of including the time-varying nature of the parameters of the model in the batch BW framework.

REFERENCES

[1] J. Proakis and C.L. Nikias, "Blind Equalization", *Proc. SPIE Adaptive Signal Processing 1991*, Vol. 1565, pp. 76-87, 1991.

[2] John Shynk et al., "A Comparative Performance Study of Several Blind Equalization Algorithms", *Proc. SPIE Adaptive Signal Processing 1991*, Vol. 1565, pp. 102-117, 1991.

[3] M. Ghosh and C.L. Weber, "Maximum Likelihood Blind Equalization", *Proc. SPIE Adaptive Signal Processing 1991*, Vol. 1565, pp. 188-195,1991.

[4] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proc. IEEE*, Vol. 77, No. 2, pp. 257-286, February 1989.

[5] R. Steele, "Mobile Radio Communications", *Pentech Press Publishers*, London 1992.

[6] T. Aulin and C.W. Sundberg, "Continuous Phase Modulations (Parts I and II)", *IEEE Trans. on Comm.*, Vol COM-29, No.3, pp. 196-225, 1981.

[7] José A.R. Fonollosa and J. Vidal, "Application of Hidden Markov Models to Blind Channel Characterization and Data Detection", *Proc. IEEE Int. Conf. Accoust. Speech and Signal Processing*, Australia, pp.185-188, April 1994.

[8] R. d'Avella, L. Moreno and M. Sant'Agostino, "An Adaptive MLSE Receiver for TDMA Digital Mobile Radio", *IEEE J. on Selected Areas in Comm.*, Vol. 7, No. 1, pp. 122-129, Jan. 1989.

[9] G. Ungerboeck, "Adaptive Maximum-Likelihood Receiver for Carrier Modulated Data Transmission Systems", *IEEE Trans. on Comm.*, Vol. COM-22, No. 5, pp. 856-864, May 1974.