A FAMILY OF NONLINEAR EQUALIZERS: SUB-OPTIMAL BAYESIAN CLASSIFIERS

J. R. Montalvão F.[(1,2)], *B. Dorizzi*[(2)] , *J. C. M.Mota*[(3)]

[(1)]Univ. Tiradentes R.Lagarto, 264, 49010-390, Aracaju-SE, Brazil jugurta.montalvao@int-evry.fr
[(2)]Inst.Nat.Télécom./EPH 9, Rue Charles Fourier 91011 Evry France Bernadette.Dorizzi@int-evry.fr
[(3)]Univ.Fed.Ceará/CT/DEE C.P.6001, Fortaleza-CE Brazil 60455-760 mota@dee.ufc.br

ABSTRACT

A family of sub-optimal Bayesian equalizers is proposed in two versions: feed-forward and decision feedback. We show that this family of equalizers provides a range of gradual choices concerning the tradeoff between equalizer complexity and symbol error rate (SER). We also point out the SER equivalence between the simplest proposed structure (the simplest equalizer of the family) and Wiener linear equalizer (or the decision feedback equalizer for the decision feedback version). Some simulations results are also presented.

1. INTRODUCTION

Intersymbol interference (ISI) is a typical problem in digital communication systems. It occurs when the communication channel has a considerable "memory", which overlaps distinct transmitted symbols. In such cases, special filters, called equalizers, are used in order to reconstruct the transmitted symbols by combating the ISI effect.

Linear equalizers have been used for long time. Their importance is associated to their low complexity and theoretical tractability. However, it has been shown [2] that the optimum equalizer is nonlinear in all realistic cases where noise is present and the channel is non-minimum phase. Indeed, considering the equalization problem as a classification one, the optimum symbol detector is the maximum a posteriori symbol-by-symbol detector MAPSD, proposed by [9]. Considering the symbol error rate (SER), the MAPSD is better than the popular Viterbi equalizer [10], which is actually an efficient implementation of the maximum likelihood sequence estimation (MLSE). However, despite its desirable performance, the MAPSD is strongly limited by its inherent structure complexity.

A possible simplification of the MAPSD consists in segmenting the full vector of observations in blocs and to proceed as if the samples in these different blocs were mutually independent. In some works (e.g. [3-5]), this simplification has been used in order to implement nonlinear equalizers with Radial Basis Function (RBF) structures. Indeed, in such approaches, the channel states are points in a finite dimensional space and then, clustering methods are applied over the channel outputs in order to reduce the complexity of the RBF equalizers. Several different methods have been proposed in order to find the cluster centers, ranging from k-means to Neural Network based algorithms[6].

Actually, in such approaches, finding the cluster centers corresponds to an implicit channel estimation. Keeping this in mind, we propose, in this work, a new method of reducing the MAPSD (per blocks) equalizer complexity using an explicit estimation of the channel parameters.

For that purpose, we perform a Bayesian classification using a multi-gaussian approximation of the probability density function of each class (one class for each element in the discrete alphabet of modulation). The quality of the approximation depends on the number of gaussians in the approximation.

The proposed structure is presented in two versions: feed-forward sub-optimal Bayesian equalizer (SBE) and a decision feedback sub-optimal Bayesian equalizer (DF-SBE). The first one is similar to a gaussian Radial Basis Function (RBF) network where each RBF is thus centered at one of the previously calculated centers and the whole input signal, structured as a random vector, passes through a nonlinear transformation corresponding to the RBF network. Equivalently, the DF-SBE uses the previously estimated symbols to outperform the SBE in low noise environments. Then, its structure is similar to a recurrent RBF.

This paper is organized as follows. In Section 2 we present the system model. The forward and decision feedback versions of the proposed family of equalizers are presented in Section 3. In Section 4 we discuss the trade-off between complexity and performance for such an approach. We present some illustrative simulation results in Section 5, for BPSK and 4-QAM modulation schemes. Finally, we summarize our major findings and outline our future work in Section 6.

2. SYSTEM MODEL

In this work, we consider a particular communication scheme where the digital data are drawn with equal probability from a finite alphabet $\{a_s : 1 \le s \le S\}$, forming an i.i.d. sequence $a(n)$ of variance σ_a^2, and the noise, $b(n)$, is additive, white and gaussian with zero mean and variance σ_b^2. The channel model is a FIR filter, whose impulse response is of length *N*. Figure 1 shows a schematic representation of such a model:

Further, it is useful to define some column vectors:

- symbol data block:

$$\mathbf{a}(n) = [a(n) \ \cdots \ a(n-d) \ \cdots \ a(n-N-M+1)]^T$$

- channel impulse response: $\mathbf{f} = [f_0 \quad f_1 \quad \cdots \quad f_{N-1}]^T$

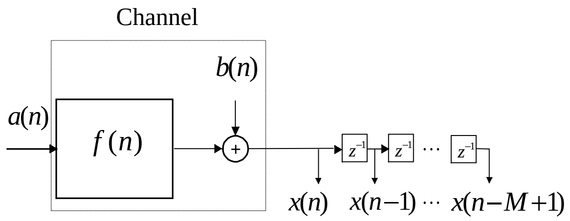

Figure 1. System Model.

- sampled noise block:

$$\mathbf{b}(n) = [b(n) \quad b(n-1) \quad \cdots \quad b(n-M+1)]^T$$

- sampled received signal:

$$\mathbf{x}(n) = [x(n) \quad x(n-1) \quad \cdots \quad x(n-M+1)]^T$$

We also define the convolution matrix:

$$\mathbf{F} = \begin{bmatrix} f_0 & & \mathbf{0} \\ \vdots & \ddots & \\ f_{N-1} & \ddots & f_0 \\ & \ddots & \vdots \\ \mathbf{0} & & f_{N-1} \end{bmatrix}_{(M+N-1)\times(M)}$$

And now we can write: $\mathbf{x}(n) = \mathbf{c}(n) + \mathbf{b}(n) = \mathbf{F}^T\mathbf{a}(n) + \mathbf{b}(n)$, where $\mathbf{c}(n)$ is the channel state vector.

3. THE PROPOSED FAMILY OF EQUALIZERS

3.1 The Feed Forward Version

Given a block of observations $\mathbf{x}(n)$, the probabilistic symbol-by-symbol algorithm (also called maximum a posteriori symbol-by-symbol (MAPSD) or full Bayesian equalizer) choose the symbol a_s which maximizes the conditional probability:

$$s = \arg\max_i \Pr(a(n-d) = a_s \mid \mathbf{x}(n)).$$

Using the Bayes rule, the conditional probability density function (pdf) associated to each symbol in the modulation alphabet, is given by:

$$\Pr(a(n-d) = a_s \mid \mathbf{x}(n)) = \frac{\Pr(\mathbf{x}(n) \mid a(n-d) = a_s)\Pr(a(n-d) = a_s)}{\Pr(\mathbf{x}(n))}$$

$$\Pr(a(n-d) = a_s \mid \mathbf{x}(n)) = \sum_{i:a(n-d)=a_s} \phi_i(\mathbf{x}) \Big/ \sum_i \phi_i(\mathbf{x})$$

where $\phi_i(\mathbf{x}) = \frac{1}{\sqrt{(2\pi\sigma_b^2)^M}} \exp\left(\frac{-(\mathbf{x}-\mathbf{c}_i)^H(\mathbf{x}-\mathbf{c}_i)}{2\sigma_b^2}\right)$ and each channel state (or center) forming the set: $\{\mathbf{c}_i : 1 \le i \le S^{M+N-1}\}$ is a conditional mean of the received signal vector:

$$\mathbf{c}_i = E\{\mathbf{x}(n) \mid \mathbf{a}(n) = \mathbf{a}_i\} = \mathbf{F}^T\mathbf{a}_i \qquad (01)$$

where $\mathbf{a}_i$ is a possible sequence of $M+N-1$ symbols. The superscript H denotes Hermitian transposition.

The full Bayesian equalizer may be summarized as:

$$\hat{a}(n-d) = \text{Dec}\left(\sum_{i=1}^{S} a_s f_{d,s}(\mathbf{x}(n))\right),$$

where $f_{d,s}(\mathbf{x}(n)) = \Pr(a(n-d) = a_s \mid \mathbf{x}(n))$.

This equalizer is optimum in the sense of minimizing the probability of a symbol error [2]. However, its large computational burden is the major shortcoming of this algorithm.

Before introducing the general formulation of the new equalizer, it's worth showing a particular case of our approach which links the Wiener equalizer (nonparametric technique) and a kind of suboptimum Bayesian equalizer (parametric technique).

In fact, if we split the S^{M+N-1} centers in S clusters, whose centers are given by:

$$\tilde{\mathbf{c}}_{d,s} = E\{\mathbf{x}(n) \mid a(n-d) = a_s\}$$

$$\tilde{\mathbf{c}}_{d,s} = \mathbf{F}^T [0 \quad 0 \quad \ldots \quad \underset{\substack{\uparrow \\ d}}{a_s} \quad \ldots \quad 0]^T \qquad (02)$$

we are able to approximate each conditional pdf by one gaussian (with ellipsoidal basis) :

$$\tilde{f}_{d,s}(\mathbf{x}(n)) = \tilde{\phi}_s(\mathbf{x}) \Big/ \sum_{i=1}^{S} \tilde{\phi}_i(\mathbf{x})$$

$$\tilde{\phi}_s(\mathbf{x}) = \exp\left(\frac{-(\mathbf{x}-\tilde{\mathbf{c}}_{d,s})^H \mathbf{R}_{x-c}^{-1} (\mathbf{x}-\tilde{\mathbf{c}}_{d,s})}{2}\right)$$

where $\mathbf{R}_{x-c} = E\{(\mathbf{x}(n) - \mathbf{f}_*^T a(n-d))(\mathbf{x}(n) - \mathbf{f}_*^T a(n-d))^H\}$ $= \sigma_a^2(\mathbf{F}^H\mathbf{F} - \mathbf{f}_*{}^H\mathbf{f}_*) + \sigma_b^2\mathbf{I}$ is the within class dispersion matrix [1] of each cluster and $\mathbf{f}_* = [f_d \quad f_{d-1} \quad \cdots \quad f_{d-M+1}]$ is an auxiliary vector.

Despite this rough pdf approximation, it's possible to show [7] that the equalizer:

$$\hat{a}(n-d) = \text{Dec}\left(\sum_{i=1}^{S} a_s \tilde{f}_{d,s}(\mathbf{x}(n))\right)$$

and the Wiener linear equalizer provide the same decision boundary in the *M*-dimensional decision space. In others words, they provide both the same SER.

In fact, we can generalize the idea of clustering centers with a number of clusters ranging from S to S^{M+N-1}. In order to do that, we calculate the S^{P+Q+1} cluster centers by the generalization of the equation 02:

$$\tilde{\mathbf{c}}^{t}_{d,(\pi,\ldots,s,\ldots,\theta)} = E\{\mathbf{x}(n) \mid a(n-d-P) = a_\pi, \ldots$$
$$\ldots, a(n-d) = a_s, \ldots, a(n-d+Q) = a_\theta\} \quad 1 \le \pi, s, \theta \le S$$

or, equivalently

$$\tilde{\mathbf{c}}^{t}_{d,(\pi,\ldots,s,\ldots,\theta)} = [0 \ \ \ldots \ \ 0 \ \ \underset{\substack{\uparrow \\ d-P}}{a_\pi} \ \ \ldots \ \ \underset{\substack{\uparrow \\ d}}{a_s} \ \ \ldots \ \ \underset{\substack{\uparrow \\ d+Q}}{a_\theta} \ \ 0 \ \ \ldots \ \ 0]\,\mathbf{F} \tag{03}$$

This calculation procedure suggests the definition of the helpful auxiliary matrix:

$$\mathbf{F}_* \overset{\Delta}{=} \begin{bmatrix} f_{d-P} & f_{d-P-1} & \cdots & f_{d-P-M+1} \\ \vdots & \vdots & \cdots & \vdots \\ f_d & f_{d-1} & \cdots & f_{d-M+1} \\ \vdots & \vdots & \cdots & \vdots \\ f_{d+Q} & f_{d+Q-1} & \cdots & f_{d+Q-M+1} \end{bmatrix}_{(P+Q+1)\times(M)}, \quad P \text{ and } Q \in \mathbb{N} \tag{04}$$

And eq. 03 can be rewritten similarly to eq. 01:

$$\tilde{\mathbf{c}}^{t}_{d,i} = \mathbf{F}_*^T \tilde{\mathbf{a}}_i \tag{05}$$

where $\tilde{\mathbf{a}}_i$ is the i-th possible sequence of $P+Q+1$ symbols.

Each approximated conditional pdf can be obtained as the sum of S^{P+Q+1} gaussians (with ellipsoidal bases) :

$$\tilde{f}_{d,s}(\mathbf{x}(n)) = \sum_{i:a(n-d)=a_s} \tilde{\phi}_i(\mathbf{x}) \Big/ \sum_{i=1}^{S} \tilde{\phi}_i(\mathbf{x})$$

where $\tilde{\phi}_i(\mathbf{x}) = \exp\left(\dfrac{-(\mathbf{x} - \tilde{\mathbf{c}}_{d,i})^H \mathbf{R}_{x-c}^{-1} (\mathbf{x} - \tilde{\mathbf{c}}_{d,i})}{2} \right)$

and $\mathbf{R}_{x-c} = E\{(\mathbf{x}(n) - \mathbf{F}_*^T \tilde{\mathbf{a}}(n))(\mathbf{x}(n) - \mathbf{F}_*^T \tilde{\mathbf{a}}(n))^H\}$

$$= \sigma_a^2 (\mathbf{F}^H \mathbf{F} - \mathbf{F}_*^H \mathbf{F}_*) + \sigma_b^2 \mathbf{I}\,.$$

Finally, the output of the feed forward Suboptimum Bayesian Equalizer (SBE) is given by:

$$\hat{a}(n-d) = \mathrm{Dec}\left(\sum_{i=1}^{S} a_s \tilde{f}_{d,s}(\mathbf{x}(n)) \right)$$

A schematic representation of the SBE is shown in Fig. 2.

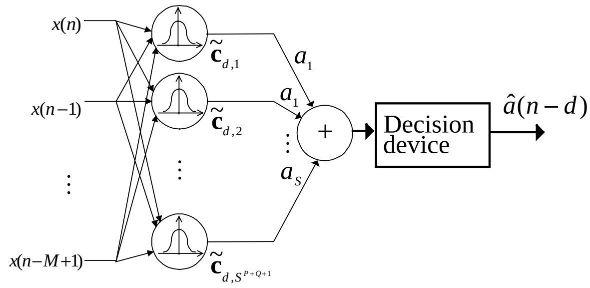

Figure 2. Schematic aspect of the SBE, where the number of gaussians is function of *P* and *Q*.

3.2 The Decision Feedback Version

The decision feedback version of the SBE (DF-SBE) uses the estimated symbols in order to improve the equalizer performance. The formulation of this version is very similar to that of the forward version, excepting the auxiliary matrix, now redefined as:

$$\mathbf{F}_* \overset{\Delta}{=} \begin{bmatrix} f_{d-P} & f_{d-P-1} & \cdots & f_{d-P-M+1} \\ \vdots & \vdots & \cdots & \vdots \\ f_d & f_{d-1} & \cdots & f_{d-M+1} \end{bmatrix}_{(P+1)\times(M)}, \quad P \text{ and } Q \in \mathbb{N} \tag{06}$$

and the clusters centers which are actualized at every time *n* by:

$$\tilde{\mathbf{c}}^{t}_{d,i}(n) = \mathbf{F}_*^T \tilde{\mathbf{a}}_i + \mathbf{F}_b^T \hat{\mathbf{a}}(n-d-1) \tag{07}$$

where $\mathbf{F}_b$ and $\hat{\mathbf{a}}(n-d-1)$ are given by:

$$\mathbf{F}_b = \begin{bmatrix} f_{d+1} & \cdots & f_{d-M+2} \\ \vdots & \ddots & \vdots \\ f_{N-1} & \ddots & \\ \mathbf{0} & & f_{N-1} \end{bmatrix}_{(M+N-d-2)\times(M)}$$

$$\hat{\mathbf{a}}(n-d-1) = [\hat{a}(n-d) \ \ \cdots \ \ \hat{a}(n-M-N+2)]^T\,.$$

Then, assuming that the estimated symbols are correct ones, the cluster scatter matrix is now given by

$$\mathbf{R}_{x-c} = E\{(\mathbf{x}(n) - \mathbf{F}_*^T \tilde{\mathbf{a}}(n) - \mathbf{F}_b^T \hat{\mathbf{a}}(n-d-1)) \cdot$$
$$(\mathbf{x}(n) - \mathbf{F}_*^T \tilde{\mathbf{a}}(n) - \mathbf{F}_b^T \hat{\mathbf{a}}(n-d-1))^H\}$$

$$\mathbf{R}_{x-c} = \sigma_a^2 (\mathbf{F}^H \mathbf{F} - (\mathbf{F}_* + \mathbf{F}_b)^H (\mathbf{F}_* + \mathbf{F}_b)) + \sigma_b^2 \mathbf{I}$$

A schematic representation of the DF-SBE is shown in Fig. 3.

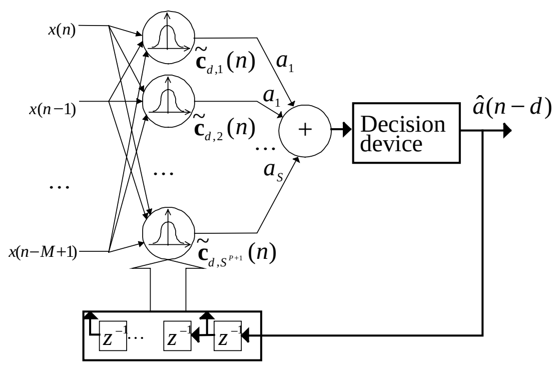

Figure 3. Schematic aspect of the DF-SBE, where the number of gaussians is function of *P*.

4. TRADE-OFF BETWEEN COMPLEXITY AND PERFORMANCE

The approach proposed in this article provides, for each channel impulse response, a family of equalizers with different levels of complexity and performance. That is, adjusting the constants *P* and *Q* (or only *P* for the DF-SBE), we can choose a suitable number of cluster centers (eqs. 05 and 07), thus adjusting the device performance.

This is specially useful when the channel impulse response (CIR) is long and the Viterbi equalizer has a prohibitive complexity. Evidently, smaller is the number of centers, worse is the equalizer performance in terms of SER. But, fortunately, we can analytically show [ ] that the simplest SBE and the Wiener equalizer (and, equivalently, the simplest DF-SBE and the classic DFE) provides the same SER.

One simple comparison of computational burden between the Viterbi equalizer and our proposition is obtained by comparing the number of states in the trellis of Viterbi and the number of centers in our approach (both the number of states and the number of centers play a similar and central role on the computational burden of each algorithm). In such a comparison, the Viterbi equalizer is $O(S^N)$ while the SBE is $O(S^{P+Q+1})$ and the DF-SBE is $O(S^{P+1})$. It's worth noticing that the simplest DF-SBE ($P=0$) corresponds to the classic DFE, which has already a SER comparable to that of the Viterbi equalizer, in high SNR [2]. Then, as the DF-SBE can outperform the DFE equalizer by choosing $P>0$, this structure can provide a very low SER, if we have a good channel estimation.

On the other hand, we have tested that the classical DFE outperforms the forward SBE for intermediate values of P and Q. Actually, the interest of the forward SBE is mostly a theoretical one. Indeed, it provides a family of analytically tractable approximations of the signal probability density function (pdf) which has been useful along the subsequent work on channel estimation.

5. SIMULATION RESULTS

The simulations are split in two parts. In the first part, concerning Figures 4 and 5, we compare some SBEs to the Wiener equalizer. Equivalently, in the second part some DF-SBEs are compared to the Viterbi equalizer and DFE (Fig. 6).

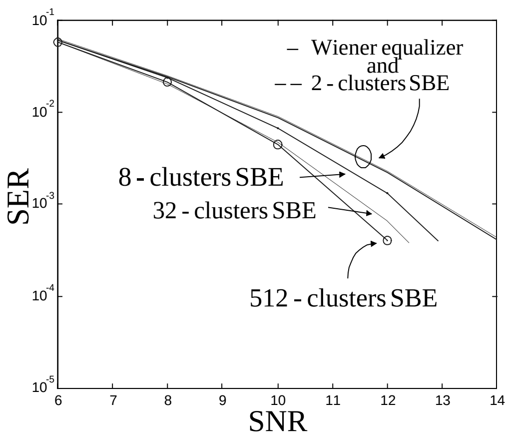

Figure 4. Performance comparison for a fixed number of taps ($M=5$), BPSK modulation, $d=4$ and channel $\mathbf{f}=[-0.21 \quad -0.50 \quad 0.72 \quad 0.36 \quad 0.21]^T$.

As a first illustration example, in Fig. 4 we show the SER performance of some SBEs. The FIR channel model is composed of 5 real coefficients, and the modulation scheme is the BPSK (-1,+1). Trials are carried out under different SNR levels (i.e. $10\log((\sigma_x^2+\sigma_b^2)/\sigma_b^2)$).

We can observe that the 2-clusters SBE and the Wiener equalizer are equivalent, as analytically foreseen. Further, greater the number of clusters, better the SBE performance.

The second example uses a 4-QAM modulation and the channel is a complex one, having two in-band spectral near-nulls. In terms of classification, a spectral null means that there is no linear separation between states of different labels, regardless the classification space dimension *M*. In such a situation, "linear" equalizers, like the Wiener one and the 4-clusters SBE, are not suitable. Actually, these two equalizers are considered in this simulation in order to show their equivalent performance. Fig. 5 shows a typical result when we vary the number of equalizer inputs *M*.

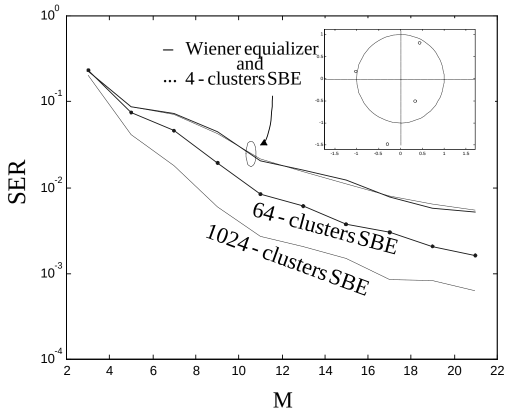

Figure 5. Performance comparison for a fixed SNR=15 and modulation 4-QAM. The channel is complex and its polynomial zeros are shown in subfigure.

Figure 6 shows the performances for a 3 coefficients complex channel, with two in-band spectral nulls. As we can observe, the DF-SBE of $O(4^2)$ provides a performance comparable to the Viterbi equalizer of $O(4^3)$.

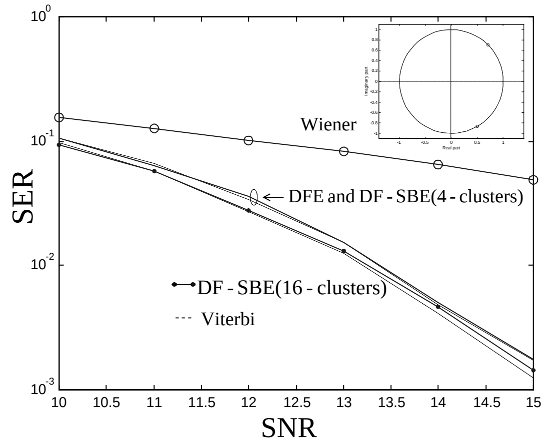

Figure 6. Performance comparison for a fixed number of equalizer inputs $M=31$, 4-QAM modulation, $d=25$. The channel is complex and its polynomial zeros are shown in subfigure.

6. CONCLUSIONS

In this paper, a suboptimum family of equalizers has been proposed based on the MAPSD. The development of such a family of equalizers was motivated by the possibility of controlling the algorithm complexity.

The SBE structure is proposed in two versions. In the forward version, the simplest equalizer of this family has a complexity equivalent to that of the Wiener linear equalizer with equal number of inputs and same decision delay. Equivalently, in the forward/backward version, the simplest equalizer of this family has a complexity and performance equivalent to that of the DFE linear equalizer with equal number of inputs and same decision delay. Some experimental results were shown as illustration of such an equivalence.

We have also shown that the equalizer complexity control is carried out by the appropriated choice of two constants, namely, P and Q (only P for the DF-SBE).

Actually, as we can observe, the DF-SBE provides a SER lower than the SBE. However, the SBE version is interesting because it gives a parametric and analytically tractable approximation of the received signal pdf. We have exploited this property in a subsequent work on channel estimation. Our first results are quite interesting.

Furthermore, it's worth noticing that each SBE is based on a channel state clustering and the center of each cluster is easily calculated by means of an auxiliary matrix (defined in eq. 05). Then, two aspects must be considered: a) the channel estimation and b) the construction of such an auxiliary matrix.

a) In fact, as we have already said, the channel estimation is the subject of our subsequent work, which uses the same ideas as the ones presented in this paper in order to provide a channel estimation.

b) The construction of the auxiliary matrix, proposed in eq. 04 and 06, subject to a fixed number of centers, is not optimal. Indeed the optimal auxiliary matrix, for a fixed number of rows, would be formed by disjoints rows of the convolution matrix. This procedure is equivalent to the problem of adjusting the RBF complexity by variable selection. Unfortunately, this point was not discussed in this paper, but it will be taken into account in future works.

7. REFERENCES

[1] DUDA, R. O. and P. E. HART (1973). *Pattern Classification and Scene Analysis*, Wiley, New York.

[2] PROAKIS, J. G. (1984). *Digital Communications,* McGraw-Hill.

[3] GIBSON, G. J., S. SIU, S. CHEN, C. F. N. COWAN and P. M. GRANT(1990). "The Application of Nonlinear Architectures to Adaptive Channel Equalization", Proc. IEEE, pp. 0649-0653.

[4] CHENG , S., B. MULGREW, P. M. GRANT (1993) "A Clustering Technique for Digital Communications Channel Equalization Using Radial Basis Function Networks", *IEEE Transaction on Neural Network* , vol. 4, NO. 4, pp. 571-579.

[5] CHNG , E., H. YANG, W. SKARBEK (1996) "Reduced Complexity Implementation of Bayesian Equalizer Using Local RBF Networks for Channel Equalization Problem", *Electronics Letters*, vol. 32, NO. 1, pp. 17-19.

[6] DORIZZI, B., J. C. MOTA, F. ALBU (1997). "A Step Towards Equalization for Radiomobile Channel: Neural Networks and Variable Selection", COST 254/IEEE.

[7] MONTALVÃO , J. R., J. C. M. MOTA, B. DORIZZI, C. C. CAVALCANTE (1998). "Bayesian Equalization Strategy: Performance versus Number of Clusters", Intern. Telecom. Symp. - ITS'98/IEEE, august, S.P., Brazil.

[8] MONTALVÃO, J. R., J. C. M. MOTA, B. DORIZZI (1998). " Some Theoretical Limits of Efficiency of Linear and Nonlinear Equalizers ", sous évaluation pour la publication à *Revista da Sociedade Brasileira de Telecomunicações.*

[9] ABEND , K. AND B. D. FRITCHMAN, "Statistical Detection for Communication Channels with Intersymbol Interference," Proc. IEEE, vol. 58, pp. 779-785, May 1970.

[10] FORNEY, G. D., "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," IEEE Trans. On Information Theory, vol. IT-18, NO. 3, pp. 363-378, May 1972.