2003 IEEE XIII Workshop on
Neural Networks for Signal Processing

UNSUPERVISED CHANNEL EQUALIZATION USING FUZZY PREDICTION-ERROR FILTERS

Rafael Ferrari[1], C. M. Panazio[2], R. R. F. Attux[1], C. C. Cavalcante[1],
L. N. de Castro[1], F. J. Von Zuben [1] and J. M. T. Romano[1]
[1]FEEC - State University of Campinas (Unicamp), Campinas-SP, Brazil
[2]CNAM - Conservatoire National des Arts et Métiers, Paris, France
E-mail: {rferrari,romisri,charles,romano}@decom.fee.unicamp.br
{lnunes,vonzuben}@dca.fee.unicamp.br, cristiano.panazio@cnam.fr

Abstract. In this work we present a new paradigm for unsupervised nonlinear equalization based on prediction-error fuzzy filters. Tests in different linear channel scenarios are carried out in order to assess the performance of the equalizer. The results show that the proposal is solid and may provide a performance close to that of a Bayesian equalizer.

INTRODUCTION

The need for optimal performance and the continuous systemic refinement are the main reasons behind the growing interest in nonlinear equalization. This interest, together with advances in the field of computational intelligence and nonlinear filtering, account for a solid research corpus, which attests the relevance of the field.

Usually, nonlinear equalizers are adapted with the aid of a pilot signal, i.e., in a supervised fashion. This is quite natural, since the usual test of structures and algorithms must be carried out in an environment as simple as possible. Furthermore, the assumption of supervised training is reasonable in some contexts and also gives rise to a more propitious scenario for optimality analysis.

However, a general nonlinear filtering paradigm should not rely on supervised learning, since a reference signal may not be available in all cases. This is the motivation behind the proposal of unsupervised equalization criteria. Although criteria based on signal statistics work well on the adaptation of linear filters, it is not certain that they will assure the correct adaptation of nonlinear filters. Ironically, this kind of problem arises exactly from the great approximation potential of nonlinear structures.

Therefore, it becomes imperative to look for unsupervised equalization criteria adequate to the problem of nonlinear filtering. In particular Cavalcante et al. [1] demonstrated that a prediction approach can be effective

0-7803-8178-5/03/$17.00

in a linear channel context. Given the theoretical solidity of this formulation, we consider it a reliable basis to build upon a unsupervised equalization paradigm, what is indeed the main objective of our work.

We propose a framework based on a fuzzy filtering prediction structure and three adaptation techniques: a modified k-means clustering algorithm, a procedure for rule generation and a classical Recursive Least Squares (RLS) algorithm. We will test our proposal with different channels and compare it with the Bayesian optimal criterion.

This paper is organized as follows. In section 2, we present a basic background on adaptive equalization and nonlinear prediction. The fuzzy structure and the adaptation procedure are presented in sections 3 and 4. Section 5 brings the results and section 6, our conclusions.

ADAPTIVE EQUALIZATION AND NONLINEAR PREDICTION

The main objective of a communication system is to assure proper information interchange between a transmitter and a receiver, both of which are interconnected by a channel, as shown in Figure 1. This picture also establishes the notation to be used in the present work. The channel is responsible for a certain level of degradation of the message transmitted, often leading to unacceptable bit error rates (BERs).

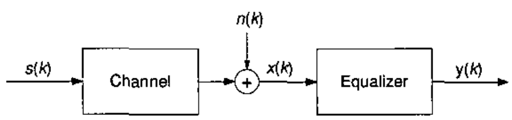

Figure 1: Simplified Model of a Communication System.

A common countermeasure to this problem is the use of an equalizer, i.e., a filter specially designed to compensate the noxious effects of the medium. Two questions then arise. 1) What filtering structure should be chosen to play the role of the equalizer? 2) How to appropriately adjust the parameters of the filter?

A linear filter is the common answer to the first question, especially due to the simple character inherent to this kind of structure. Linear filters also have a simple mathematical treatment and require a reasonably low computational cost to implement. Nevertheless, linearity implies serious performance limitations, what is clearer when nonlinear channels are at the order of the day, but can also be pronounced in the classical linear scenario.

Due to these features, there has been a growing interest in nonlinear equalizers. This tendency can be verified by a literature review on signal processing in the last two decades. The simultaneous advance in the field of computational intelligence is also responsible for the vast interest in nonlinear and adaptive filtering. Research on neural networks, fuzzy filters and others nonlinear structures experienced a major increase, bringing together several scientific communities around a highly interdisciplinary set of models and tools.

The success of these devices has several reasons, and one is of particular importance to nonlinear filtering: most of these tools are capable of approximating, with an arbitrary precision, any continuous nonlinear function defined on a compact set. That is, they are universal function approximators. As equalization can be viewed as an inverse mapping problem, nonlinear universal approximators are strong candidates to be used as general models of equalizers.

The second question is related to how to adjust the parameters of the filter. The usual answer is "through a well-designed equalization criterion". Criteria are built to guide the choice of the parameters, since they are essentially mathematical statements of a chosen equalization objective. When a training signal is available (supervised training), the goal can be made as: "make the equalizer output as close as possible to the desired signal". This is exactly the rationale behind the minimum mean square error (MMSE) criterion

$$J_{MMSE} = E\left[s(k-d) - x(k)\right]^2 \tag{1}$$

where d is the equalization delay.

However, if reference samples of a desired signal are not available a new criterion must be sought. Typical sources of information are the statistics of the transmitted signal, which are the core of all unsupervised technique. There are several examples of members of this class, ranging from the simple decision-directed criterion to the refined Shalvi-Weinstein framework [5].

Unfortunately, these criteria were designed to adapt linear filters, and their applicability to the general problem of equalization is still an open question. However, there is another unsupervised approach that can safely adapt a general filtering device in some relevant situations: the prediction paradigm [1].

Nonlinear Prediction and Equalization

A common task in signal processing is to predict future samples of an information signal from past ones. In the equalization context, a predictor is the key to a scheme of redundancy removal [1], which can be, as will be shown later, a valuable equalization approach.

This idea of prediction can be translated to mathematical notation as follows:

$$x_p(k) = f_p\left[\boldsymbol{\chi}(k-1)\right] \tag{2}$$

where $x_p(k)$ is the predicted signal, $\boldsymbol{\chi}(k-1)$ is a set of past samples, and $f_p[\cdot]$ corresponds to the mapping performed by the predictor. The prediction-error is defined as:

$$e_p(k) = x(k) - x_p(k) = x(k) - f_p\left[\boldsymbol{\chi}(k-1)\right] \tag{3}$$

The design of a predictor involves the minimization of the expected value of the square of the prediction error. This rule is intuitive and efficient. Within the equalization domain, assume, as a rule from now on, that a linear communication channel is being analyzed. Its input-ouput relation is given by

$$x(k) = \sum_{i=0}^{N_c} h(i)s(k-i) \tag{4}$$

where $h(k)$ is the channel impulse response, and N_c+1 is its length. No noise is assumed to exist yet.

To understand the link between prediction and equalization, it is necessary to consider (2) and (4). They reveal that $x_p(k)$ will be a function of $x(k-1), x(k-2)$ and so on. This implies that, given the channel model, it will also be a function of $s(k-1), s(k-2)$ and so forth. Nevertheless, Equation (3) reveals that the prediction-error depends on $x(k)$ as well. From (4), it is clear that $x(k)$ is determined by $s(k)$ as well; a process that cannot be part of $f_p[\chi(k-1)]$. This implies that in many cases of interest there will always be a residual prediction error. In the ideal case, it will be simply a scaled version of $s(k)$.

This means that the objective of a prediction-error filter and a MMSE filter with zero equalization delay will be, to some extent, equivalent. In other words, a good predictor may produce, in an unsupervised way, an estimate of the transmitted signal $s(k)$.

Such conclusion is well known from the linear prediction theory, with the immediate restriction to the minimum-phase channels. However, Cavalcante et al. [1] showed that one could overcome this limitation by using a nonlinear predictor. This kind of structure can fairly approximate an ideal prediction-error filter, thus providing a solid basis for unsupervised nonlinear equalization of linear channels. One should not underestimate the relevance of this result, given the major differences between linear and nonlinear filters when an unsupervised equalization problem is faced.

The solidity of the prediction approach is the main reason behind its choice to compose our paradigm of unsupervised nonlinear equalization, which we shall discuss in section 3.

Bayesian Equalization

When the channel is modeled as a FIR filter and the transmitted signal belongs to a finite alphabet, the received samples will also belong, in the absence of noise, to a finite set. As can be deduced from (4), each possible value of a general input vector $\mathbf{x}(k) = [\ x(k) \ \ x(k-1) \ \ \cdots \ \ x(k-m+1)\ ]^T$, where $m-1$ is the order of the equalizer, is associated with a sequence $s(k),\ s(k-1),\ \cdots,\ s(k-Nc-m+1)$. This means that, given a predetermined equalization delay d, there will be a "$s(k-d)$" label associated with each possible value of $\mathbf{x}(k)$. The task of an equalizer, in this context, will be to separate the classes of input vectors that correspond to each symbol of the transmitted signal alphabet.

Assume that the transmitted signal belongs to a binary (± 1) alphabet and has independent and identically distributed samples, that the channel obeys the general model of (4), and that the noise $n(k)$ is a non-zero additive white Gaussian noise (AWGN). To obtain the optimal decision boundary, the usual procedure is to seek the minimum probability of error, in accordance with Bayes' theory. From this criterion the following optimal solution can be obtained [4]:

$$\mathcal{F}[\mathbf{x}(k)] = \sum_{\mathbf{x}_j \in X^+} \exp\left(-\frac{\|\mathbf{x}(k) - \mathbf{x}_j\|}{2\sigma_n^2}\right) - \sum_{\mathbf{x}_i \in X^-} \exp\left(-\frac{\|\mathbf{x}(k) - \mathbf{x}_i\|}{2\sigma_n^2}\right) \quad (5)$$

where $\mathbf{x}_j$ and $\mathbf{x}_i$ are the classes of states associated with the $+1$ and -1 symbols respectively, and X^+ and X^- are the sets containing all these input vectors. The variance of the zero-mean Gaussian noise is σ_n^2.

An equalizer with a decision function equals to the one presented in (5) is called a Bayesian equalizer, i.e., one optimal in the minimum probability of error sense. It is important to note that the decision function is usually nonlinear. This implies, as outlined before, that the optimal equalization device in a linear scenario is commonly a nonlinear filter, reinforcing the general interest in this class of structures.

The Bayesian paradigm permeates the theoretical assumptions behind our proposal, which we analyze in the next section.

THE FUZZY PREDICTION-ERROR EQUALIZER (FPEE)

Fuzzy filters are nonlinear devices that process information by means of a specific set of fuzzy "IF-THEN" rules, which may be adapted in real time. Such structures have been studied and applied with success to several engineering problems.

In the context of digital signal processing, fuzzy filters are of particular interest because, as shown by Patra [4], they can play the role of a Bayesian equalizer, what implies the possibility of achieving optimal performance. This important feature, together with the prediction potential of fuzzy filters led us to choose this structure as the basis of our paradigm.

Our aim, as discussed before, is to design a fuzzy predictor capable of providing good estimates of the output of a channel modeled as in (4), with AWGN. Its role is to produce an estimate of $x(k)$ from the channel output samples $x(k-1),\ x(k-2),\ \cdots,\ x(k-m+1)$, which correspond to vector $\boldsymbol{\chi}(k-1)$.

In order to implement a fuzzy predictor, the first step is to define the fuzzy sets of the input space. Let us define $M = 2^{Nc+1}$ fuzzy sets for each input $x(k-i),\ 0 \leq i \leq m-1$, where M is the number of noise-free scalar channel output states $C_j,\ 1 \leq j \leq M$. Inspired by the fuzzy implementation of the optimal Bayesian equalizer [4], these fuzzy sets are represented by Gaussian membership functions with centers C_j and spread values σ_n,

$$\psi_i^j\left[x(k-i)\right] = \exp\left[-\frac{(x(k-i)-C_j)^2}{2\sigma_n^2}\right],\ 1 \le j \le M \text{ and } 0 \le i \le m-1 \quad (6)$$

The rule base of the fuzzy predictor is formed by rules of the following type:

$R^{(j1,j2,\ldots,jm)}$:If $x(k-1)$ is ψ_1^{j2} and ... and $x(k-m+1)$ is ψ_{m-1}^{jm} then $x(k)$ is ψ_0^{j1}

where $1 \le ji \le M$. It is not necessary to use all the possible combinations of i and j to construct the rule base of the predictor. Only those combinations associated with the noise-free channel output states are of interest. Consequently, the number of rules is equal to the number of noise-free channel output states, $N_s = 2^{N_c+m}$. Considering product inference and center of gravity defuzzification, the fuzzy predictor output is,

$$x_p(k) = f_p\left[\boldsymbol{\chi}(k-1)\right] = \frac{\sum_{l=1}^{N_s} w_l \left\{\prod_{i=1}^{m-1} \phi_{li}\left[x(k-i)\right]\right\}}{\sum_{l=1}^{N_s}\left\{\prod_{i=1}^{m-1} \phi_{li}\left[x(k-i)\right]\right\}} \quad (7)$$

where w_l is the center of the membership function related to the THEN part of the rule l and $\phi_{li}\left[x(k-i)\right], 1 \le i \le m-1$, denote the ith membership function associated to the same rule.

We can note that half of the rules are conflicting rules, i.e., rules that have the same IF part but a different THEN part. This occurs because the input vector of the predictor represents only 2^{N_c+m-1} noise-free states. For each $\boldsymbol{\chi}(k-1)$ state, there are two possible values for the noise-free value of $x(k)$, each one associated with a possible transmitted symbol, $+1$ or -1. Therefore, half the rules can be suppressed without loss of performance if the weights w_l are properly adjusted. It is possible to show that the values of the weights that minimize the prediction-error are the mean between the centers of the membership function of the THEN part of the conflicting rules. Consequently, the output of the FPEE, i.e., the prediction-error, is,

$$e_p(k) = x(k) - \frac{\sum_{l=1}^{N_s/2} w_l \left\{\prod_{i=1}^{m-1} \phi_{li}\left[x(k-i)\right]\right\}}{\sum_{l=1}^{N_s/2}\left\{\prod_{i=1}^{m-1} \phi_{li}\left[x(k-i)\right]\right\}} \quad (8)$$

The prediction-error $e_p(k)$ is then passed through a slicer in order to recover the transmitted symbol $s(k)$. A block diagram that summarizes the proposed FPEE is shown in Figure 2.

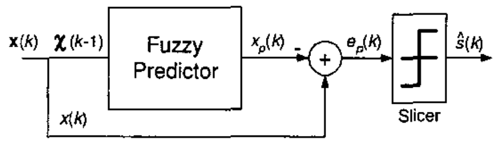

Figure 2: Fuzzy Prediction-Error Equalizer block diagram.

FPEE Adaptation Procedure

To implement the FPEE, it is necessary to estimate the N_s noise-free channel states, the noise variance σ_n^2, and the weights w_l. The first step is the estimation of the scalar channel states C_j and the noise variance. This is made by using an unsupervised clustering technique, the Enhanced k-means Algorithm [3]. The symmetry of the values of the channel states is used to speed up the convergence of the clustering algorithm [4], because, due to this feature, only half of the states need to be estimated. The initial values of the centers are uniformly distributed in the interval $[-1, +1]$, and the initial variances are all set to a small value. The k-means algorithm requires knowledge of the number of scalar channel states, which can be obtained from an autocorrelation-based estimation of the channel order [2].

The second step is the generation of the rule base using the algorithm proposed in [7]. Given the equalizer input vector $\mathbf{x}(k) = [\ x(k) \quad x(k-1) \ \ldots \ x(k-m+1)\ ]^T$, for each membership function $\psi_i^j\ [x(k-i)]$ we calculate the membership value for all vector components. The rule is obtained by assigning to each value of the input vector the fuzzy set corresponding to the membership function that achieved the maximum degree of membership. This procedure is repeated for all channel output samples in the set used in the estimation of the scalar channel output states. The rule base obtained contains conflicting rules, as discussed previously. Each pair of conflicting rules is substituted by a single rule with the same IF part and the center of the membership function of the THEN part is set to the mean value between the centers of conflicting THEN parts. These centers are the values of the weights w_l.

The estimation of the channel states and the noise statistics can be imperfect. In order to refine them, the last step is to employ a classical weight update procedure based on the well-established RLS algorithm [6].

PERFORMANCE EVALUATION AND DISCUSSION

This section investigates the performance of the proposed FPEE in terms of BER for a minimum and a non-minimum phase channel. The experiments were performed until either 3000 errors were observed or 10^6 symbols were transmitted.

The first channel used in the simulations was $h_1(z) = 0.8354+0.5012z^{-1}+0.2256z^{-2}$. This is a minimum phase channel with 2 zeros situated at $z_1 =$

-0.9 and $z_2 = 0.3$. The scalar noise free channel states are located at $\pm 1.5608, \pm 1.1096, \pm 0.5584$ and $\pm\ 0.1072$. Figure 3 shows the BER performance for the FPEE with estimated channel states, the FPEE with exact channel states, the linear prediction-error equalizer (LPEE) and the Bayesian equalizer. For all equalizers, the number of inputs m was set to 3. The values for the coefficients of the LPEE were set to the MMSE solution and were fixed over the experiments. The Bayesian equalizer equalization delay was set to $d = 0$. The BER curve for the trained version of the FPEE was obtained in the following way: the channel scalar states and the noise statistics were evaluated using the unsupervised k-means clustering algorithm with 500 samples. This same set of samples was used to generate the rule base of the fuzzy predictor. After the equalizer was constructed, the predictor weights were trained using the RLS algorithm with 500 samples. This procedure was repeated for each signal to noise ratio (SNR) value. The resulting BER curve represents the average of 30 experiments.

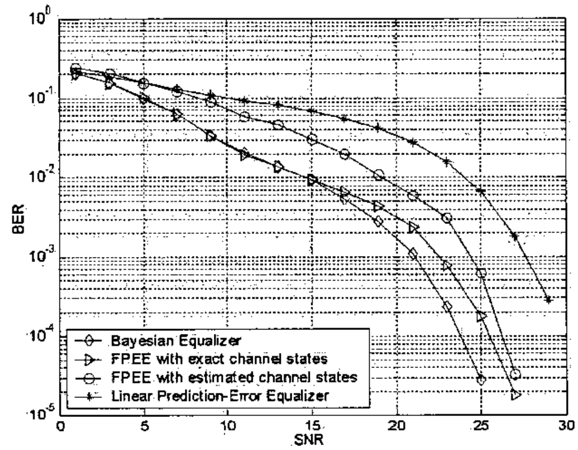

Figure 3: BER performance for FPEE with estimated channel states, FPEE with exact channel states, LPEE and Bayesian equalizer for channel $h_1(z)$, $m = 3$ and $d = 0$.

From the equalizer BER curves in Figure 3, it can be seen that the performance of the FPEE with exact channel states is close to the Bayesian equalizer. The trained version of FPEE suffers from performance degradation, mainly for low SNR values. This is because, as observed during the simulations, the estimation of the scalar channel states by the clustering algorithm and the construction of the rule base are corrupted by noise, leading to incorrect values of scalar states and rules. Nevertheless, for high values of SNR, the performance of the trained version tends to converge to that of the FPEE with exact channel states. All the equalizers outperform the LPEE.

The channel used in the next experiment was $h_2(z) = 0.5632 - 0.7322z^{-1} - 0.3830z^{-2}$. This is a non-minimum phase channel with 2 zeros located at $z_1 = 1.7$ and $z_2 = -0.4$. The scalar states are situated at $\pm 1.6784, \pm 0.9124, \pm 0.552$

and ±0.214. The procedure to generate the BER curves was the same of the previous channel.

From the simulation results depicted in Figure 4, it can be observed that for this channel the performance of the FPEE with $m = 3$ is not so close to Bayesian equalizer of the same order and $d = 0$ as it was for the previous channel. This may be because the predictor, which represents the nonlinear part of the FPEE, uses $m - 1$ inputs, while the Bayesian equalizer uses m. This difference makes the FPEE less flexible than the Bayesian equalizer of the same order. For non-minimum phase channels, which needs highly nonlinear equalizer decision boundaries for equalization delay $d = 0$, the performance degradation due to this difference is emphasized. This problem can be minimized increasing the FPEE order, as can be observed from Figure 4. The FPEE with exact noise free channel states and $m = 4$ performance is very close to Bayesian equalizer with $m = 3$. The behavior of the trained version of the FPEE with $m = 3$ is analog to that of the previous channel.

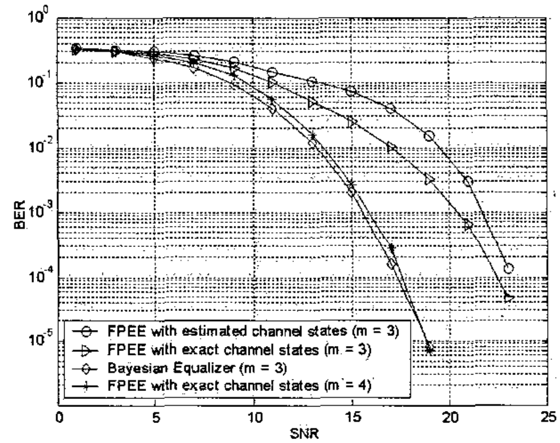

Figure 4: BER performance for FPEE with estimated channel states, FPEE with exact channel states, LPEE and Bayesian equalizer for channel $h_2(z)$ and $d = 0$.

Further experiments demonstrated that the trained version of the FPEE experiences severe degradation when there are scalar channel states very close to each other. We consider that clustering and rule generation difficulties are the main reasons for this outcome.

CONCLUSION

This work introduced a new paradigm for unsupervised nonlinear equalization based on a fuzzy structure and a prediction criterion. The proposal was presented in connection with its theoretical basis and with its motivations, considered by us very relevant, given the incipience of the field of blind nonlinear equalization.

We chose two representative channel models to assess the performance of the proposed paradigm. In the minimum-phase scenario, the ideal FPEE fairly emulated the Bayesian equalizer, though the adapted FPEE was somewhat inferior, due to clustering and rule generation problems. The nonminimum phase scenario was harder, as one would expect given the higher nonlinear character inherent to the problem of zero-delay equalization in this case. Even in this case though, the nonlinear predictor performed quite well despite the same learning problems experienced in the previous case.

We must not forget that a nonlinear prediction-error equalizer with m inputs operates effectively only on $m - 1$ samples, because of the structure inherent to a prediction-error filter. The addition of another input led to a performance almost equal to that of a Bayesian equalizer.

From these results, it is possible to conclude that the combination of a fuzzy structure, well-established learning procedures, and a prediction criterion form a solid basis for unsupervised nonlinear equalization. Our next goals are to refine both the clustering method and the rule generation procedure, and to consider in more detail other approaches to the unsupervised nonlinear problem.

ACKNOWLEDGEMENTS

The authors wish to thank FAPESP and CNPq for their financial support.

REFERENCES

[1] C. C. Cavalcante, J. R. Montalvão, Filho, B. Dorizzi and J. C. M. Mota, "A Neural Predictor for Blind Equalization in Digital Communication: Is It Plausible?" in **Proceedings of IEEE Neural Networks for Signal Processing**, Sydney, Australia, December 2000, pp. 736–745.

[2] S. Chen, B. Mulgrew and P. Grant, "A Clustering Technique for Digital Communications Channel Equalization Using Radial Basis Function Networks," **IEEE Trans. on Neural Networks**, vol. 4, no. 4, pp. 570–579, April 1993.

[3] C. Chinrungrueng and C. Sequin, "Optimal Adaptive K-Means Algorithm with Dynamic Adjustment of Learning Rate," **IEEE Trans. on Neural Networks**, vol. 6, no. 1, pp. 157–169, January 1995.

[4] S. K. Patra, **Development of Fuzzy System Based Channel Equalizers**, Ph.D. thesis, Department of Electronics and Electrical Engineering, Edinburgh University, UK, 1998.

[5] O. Shalvi and E. Weinstein, "New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels)," **IEEE Trans. on Information Theory**, *vol. 36, no. 2, pp. 312–321, March 1990.*

[6] L. Wang and J. Mendel, "Fuzzy Adaptive Filters, with Application to Nonlinear Channel Equalization," **IEEE Trans. on Fuzzy Systems**, vol. 1, no. 3, pp. 161–170, March 1993.

[7] L. Wang and J. M. Mendel, "Generating Fuzzy Rules by Learning from Examples," **IEEE Trans. on Systems, Man, and Cybernetics**, vol. 22, no. 6, pp. 1414–1427, June 1992.