

THE APPLICATION OF NONLINEAR ARCHITECTURES TO ADAPTIVE CHANNEL EQUALISATION

G. J. Gibson[†], Sammy Siu, S. Chen, C. F. N. Cowan, and P. M. Grant.

Department of Electrical Engineering, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JL, Scotland.

[†] Plessey Research Roke Manor, Romsey, Hants., SO51 0ZN, England.

ABSTRACT

This paper investigates the potential value of applying nonlinear architectures to the adaptive equalisation of a bipolar signal passed through a dispersive channel in the presence of additive noise. The problem is first described as a geometrical classification problem and consideration is given to the factors affecting the complexity of the problem. In particular, it is shown that some degree of non-linearity is desirable in the equaliser structure when signal to noise ratios are poor. The paper then proceeds by describing two nonlinear architectures which might be considered as adaptive equalisers, before evaluating their performance in this role by means of computer simulation. From this, the advantage which these structures enjoy over their linear counterparts is readily discerned.

1. INTRODUCTION.

In this paper we investigate the application of nonlinear structures as adaptive channel equalisers and demonstrate the advantage which they can offer over the linear transversal equaliser (LTE) [1], especially in the high noise environment.

The system which we shall consider throughout this paper is that depicted in Figure 1. A random sequence x_i is passed through a linear dispersive channel of finite impulse response (FIR), which is modelled by a FIR filter with response function $a_0 + a_1z^{-1} + \dots + a_kz^{-k}$, where the coefficients a_j , $0 \leq j \leq k$, are all real valued and a_0, a_k are nonzero, to produce a sequence of outputs, y_i where $y_i = \sum_{j=0}^k a_j x_{i-j}$. A term, n_i , which represents additive noise in the system, is then added to each y_i to produce an observable sequence \hat{y}_i . The problem to be considered is that of utilising the information represented by the observed channel outputs $\hat{y}_i, \hat{y}_{i-1}, \dots, \hat{y}_{i-m+1}$ to produce an estimate of the input symbol, x_{i-d} . A device which performs this function is known as an equaliser. The integers m and d are known as the order and the delay of the equaliser respectively. Throughout, the input samples are chosen from $\{-1, 1\}$ with equal probability and are assumed to be independent of one another. The additive noise samples, n_i , are chosen independently from a Gaussian distribution with mean 0 and variance

n^2 . The above system has been used to model a variety of communications systems such as the HF communications channel [2].

Traditionally, channel equalisation is considered as a deconvolution problem in which the aim is to construct an equaliser such that the impulse response of the channel/equaliser combination is as close to z^{-d} as possible. By contrast, the approach adopted in this paper is to utilise the fact that the transmitted signal is bipolar in nature and to design an equaliser which will reconstruct such a signal as accurately as possible. When viewed in this light, the problem can be viewed as a geometric spatial decision problem and some of the shortcomings of the LTE readily discerned.

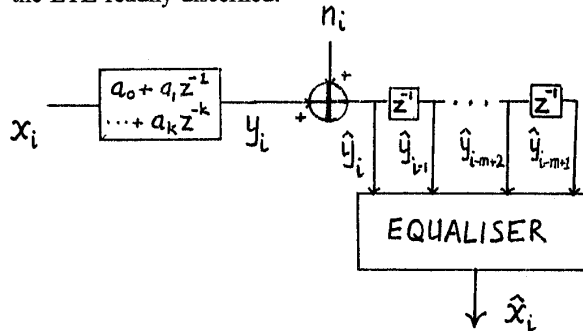


Figure 1. Schematic of data transmission system.

2. GEOMETRIC FORMULATION OF THE EQUALISATION PROBLEM

We shall continue to use the notation and definitions introduced in the previous section. In addition we make the following definitions. For the system depicted in Fig.1, let

$$P_{(m,d)}(-1) = \{ \underline{y}_i \in R^m \mid x_{i-d} = -1 \},$$

where \underline{y}_i denotes the vector of channel outputs $(y_i, y_{i-1}, \dots, y_{i-m+1})$, with $P_{(m,d)}(1)$ equivalently defined. Thus $P_{(m,d)}(1)$ and $P_{(m,d)}(-1)$ represent the sets of possible channel output vectors (y_i, \dots, y_{i-m+1}) which can be produced from sequences of channel inputs, x_i, x_{i-1}, \dots , which have $x_{i-d} = 1$ and $x_{i-d} = -1$ respectively. The observed channel output vectors, $\hat{\underline{y}}_i$, represent elements of $P_{(m,d)}(1)$ or $P_{(m,d)}(-1)$ which are corrupted in each component independently by Gaussian noise, and the task of the equaliser is to decide whether an observed output vector, $\hat{\underline{y}}_i$ represents the noise corr-

312.8.1.

CH2829-0/90/0000-0649 \$1.00 © 1990 IEEE

0649

option of an element of $P_{(m,d)}(1)$ or $P_{(m,d)}(-1)$. An equaliser must therefore produce some function $g: \mathbb{R}^m \rightarrow \{-1, 1\}$, where $\hat{x}_i = g(\hat{y}_i)$.

From Figure 1 we can readily discern the main factors which will contribute to the equalisation problem, these being the response function of the channel, $a(z)$, and the power of the additive noise, n^2 . Given a priori knowledge of these variables it is possible to specify the equaliser of order m and delay d which is optimal in the sense that it minimises the probability of a wrong decision in the estimation of the transmitted symbol x_{i-d} as follows. Let f_1 be the probability density function of the observed channel output vectors

$$\{\hat{y}_i \in \mathbb{R}^m \mid \underline{y}_i \in P_{(m,d)}(1)\}$$

and let f_{-1} be similarly defined. Now consider the equaliser whose output is the function defined by

$$g(\hat{y}_i) = \text{sgn}(f_1(\hat{y}_i) - f_{-1}(\hat{y}_i)),$$

where $\text{sgn}(x) = 1$ if $x \geq 0$ and -1 otherwise. It can be proved that this equaliser achieves the least possible bit error rate (BER) for its order. The *decision boundary* of the optimal equaliser consists of the set of points

$$\{\underline{y} \in \mathbb{R}^m \mid f_1(\underline{y}) = f_{-1}(\underline{y})\}$$

and represents the topological boundary of the closed set $g^{-1}(1)$.

Figure 2 illustrates the sets $P_{(2,0)}(1)$ and $P_{(2,0)}(-1)$, whose elements are represented by the symbols \bullet and \times respectively, for the channel with response function $1.0 + 0.5z^{-1}$, along with the optimal decision boundary for several different values of additive noise power, n^2 .

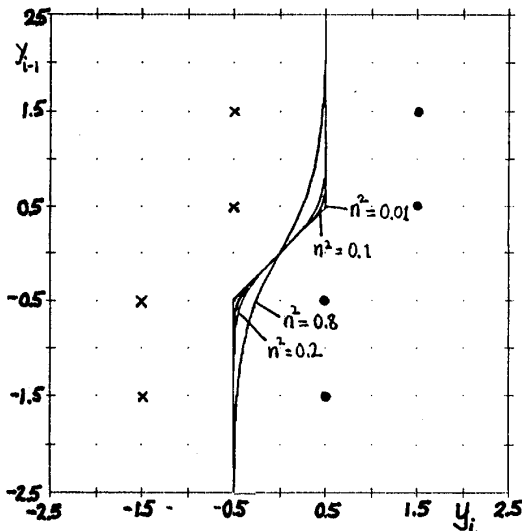
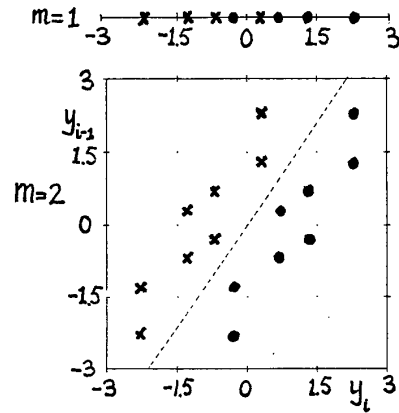


Figure 2. Channel output points and optimal decision boundaries.

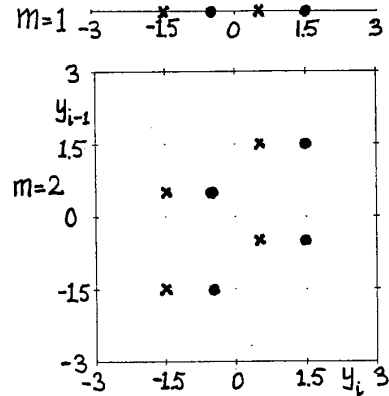
We observe that for very low levels of additive noise the decision boundary has a piecewise linear appearance, becoming progressively smoother as the noise power increases. Furthermore we remark that the optimal decision boundary is certainly nonlinear. This points us to one of the shortcomings of the LTE which necessarily forms decision boundaries which are hyperplanar and

therefore different from the optimal boundary, leading to a significantly suboptimal bit-error rate in the high noise environment.

Another shortcoming of the LTE is that it requires the sets $P_{(m,d)}(-)$ to be linearly separable if it is to successfully reconstruct the transmitted sequence. This will not in general be true, and for $d=0$ it will be true only if the channel is minimum phase (all roots of $z^k a(z)$ must lie strictly within the unit disc in the complex plane). This is illustrated for examples of minimum phase and nonminimum phase channels in Figure 3.



(a) $a(z) = 1.0 + 0.8z^{-1} + 0.5z^{-2}$ (minimum phase).



(b) $a(z) = 0.5 + 1.0z^{-1}$ (nonminimum phase).

Figure 3. Channel output points

Furthermore, in the high noise situation the LTE suffers from the problem of noise enhancement. As we allow the order of the equaliser to increase, so we increase the total power of the noise on the equaliser input, and this tends to diminish any advantage gained by increasing the equaliser order.

Figure 4 quantifies this effect by displaying the relationship between the Wiener solution of order m , to which an LTE trained by the LMS algorithm should approximate after convergence, and the resultant BER in a variety of noise environments for the equalisation of

the above channel. In the high noise environment it is clear that there is little to be gained in terms of BER by increasing the order of an LTE beyond 4 or 5. At this point we recall from Figure 2 the nonlinear nature of the decision boundary of the theoretically optimal equaliser of order 2. When this is taken into consideration along with the curves displayed in Figure 4, this provides some justification for considering equalisers of low order which have some nonlinear decision making ability incorporated into their design, when noise conditions are adverse. In the following section we describe two structures with this capability.

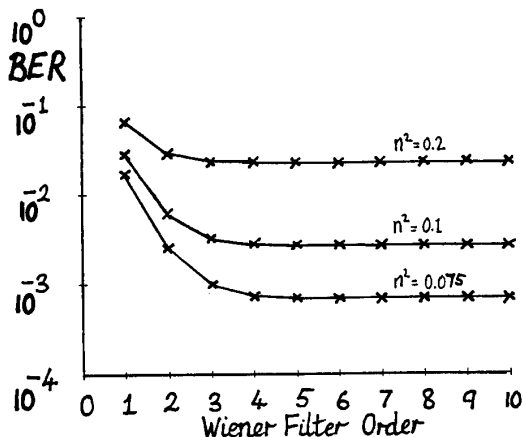


Figure 4. Relationship between Wiener filter order and BER.

3. NONLINEAR ARCHITECTURES

In this section we give a brief description of the architecture and capabilities of the multilayer perceptron (MLP) [3] and the Volterra series classifier (VSC), the structures which we shall consider as the candidates for our alternative equaliser design.

The basic building block of the multilayer perceptron is the single neuron or node which we depict in Figure 5 below. A node receives a number of real inputs x_1, \dots, x_n , say, which are then multiplied by a set of weights w_1, \dots, w_n and the resultant values are summed. To this weighted sum of inputs is added a constant θ , known as the node threshold and the output of the node is obtained by evaluating a nonlinear function, f , of the total. In this paper we restrict our attention to perceptrons where the node activation function, f , is defined by

$$f(x) = (1 - e^{-x}) / (1 + e^{-x})$$

the graph of which is shown in Figure 5.

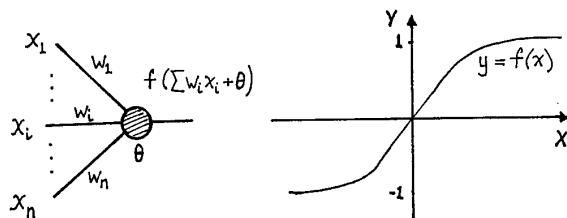


Figure 5. Perceptron node structure and activation function.

In the multilayer perceptron a number of nodes of the type described above are arranged in layers, as depicted in Fig.6. A multi-dimensional input is passed to each node in the first layer. The outputs of the first layer nodes then become inputs to the nodes in the second layer, and so on. The output of the network is therefore the outputs of the nodes lying in the final layer. Thus, weighted connections exist from a node to every node in the succeeding layer, but no connections exist between nodes in the same layer. We shall describe the architecture of a perceptron by a sequence of integers $n_0 - n_1 - \dots - n_j$ where n_0 is the dimension of the input to the network, and the number of nodes in each layer, ordered from input to output, is n_1, \dots, n_j . In this notation, the perceptron therefore produces a nonlinear mapping, $g: R^{n_0} \rightarrow R^{n_j}$.

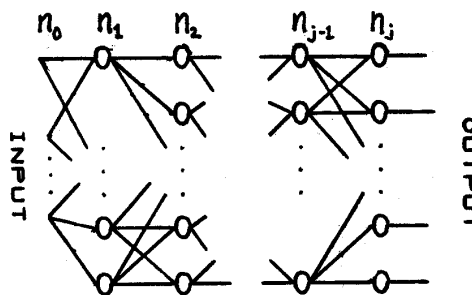


Figure 6. Multilayer perceptron structure.

In this paper we shall concern ourselves with perceptrons which have a single node in the output layer and which therefore produce a mapping $g: R^{n_0} \rightarrow R$. For such a perceptron, we define its *decision region* to be the set of points

$$\{x \in R^{n_0} \mid g(x) \geq 0\}.$$

Perceptrons which have three layers are essentially capable of forming any desired decision region and it is this architecture which we shall consider as a possible candidate for our nonlinear equaliser. When required to learn a task adaptively, perceptrons can be trained by means of the *back-propagation* (BP) algorithm [4]. This a simple stochastic gradient descent algorithm whose mode of operation is as follows. At the i^{th} iteration a training sample x_i is presented to the perceptron and the resultant output, $g(x_i)$, compared with a desired output, $d(x_i)$ (which will always be 1 or -1 in our application), to produce an error, e_i . The weights and thresholds of the nodes in the network are then updated in such a manner as to decrease the square of the error e_i , according to the following equations.

$$(w_{i+1}, \theta_{i+1}) = (w_i, \theta_i) + \Delta_i$$

$$\Delta_i = -\beta \frac{\partial (e_i^2)}{\partial (w_i, \theta_i)} + \alpha \Delta_{i-1}$$

where β is the adaptive gain and α is a momentum parameter which "smooths out" high frequency variations in the weight/threshold vector during training.

The alternative nonlinear architecture incorporates polynomial nonlinearities in its structure and we shall refer to it as a Volterra Series Classifier (VSC). Its

mode of operation is as follows. Given an input, $x \in R^t$, the VSC produces an output

$$g(x) = f(p_{\hat{g}}(x))$$

where f is the function defined above and $p_{\hat{g}}(x)$ is a polynomial in the components of x of fixed degree, t . We shall refer to this architecture as an (s, t) VSC. The vector \hat{g} represents the coefficients of the polynomial and has dimension equal to the binomial coefficient $\binom{t+s}{s}$. Like the multilayer perceptron the VSC can be trained by performing gradient descent on the weight vector \hat{g} . We can similarly define the decision region of the VSC to be the set of inputs resulting in a positive output, just as was done for the MLP.

4. SIMULATION RESULTS

In a previous section we described the manner in which the performance of an LTE of order m degrades under certain conditions on the channel response function, $a(z)$, and the power of the additive noise, n^2 . In the first case, we saw that the sets $P_{(m,0)}(1)$ and $P_{(m,0)}(-1)$ are never linearly separable if the channel is nonminimum phase. We also considered the problem of noise enhancement, which restricts the length of LTE which can effectively be employed. We now present simulations which illustrate the potential advantages which nonlinear structures might offer over the LTE. In all the simulations presented the system depicted in Figure 7 was employed.

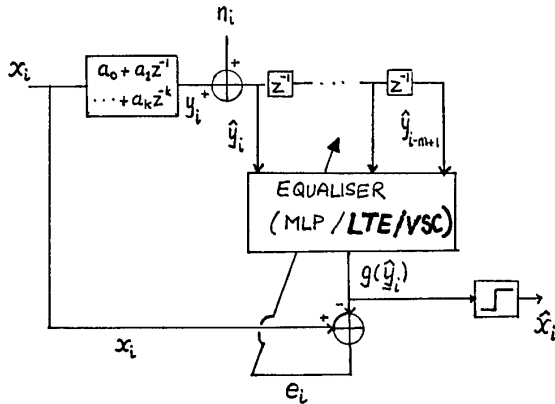
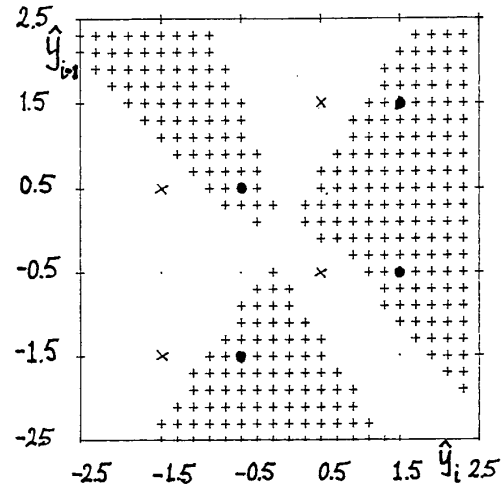


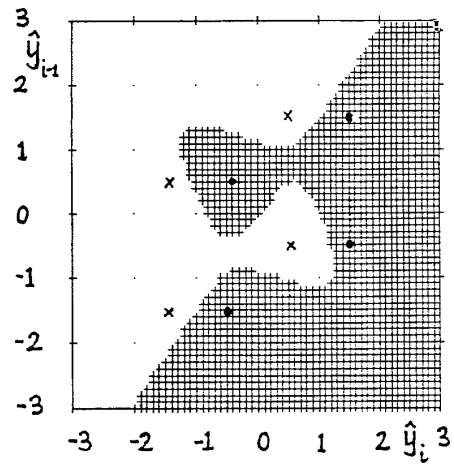
Figure 7. System employed for simulations.

The first simulation results presented demonstrate the performance MLP and VSC employed to equalise the channel with response function $0.5 + 1.0z^{-1}$, for which the sets $P_{(m,0)}(-)$, $m=1,2$ were depicted in Figure 3. As this channel is nonminimum phase, a linear equaliser is incapable of reconstructing the input sequence (without the introduction of some delay). In Fig.8 we show the decision region formed by a 2-6-4-1 perceptron after a training sequence of 1000 samples, along with the decision region of (2,5) VSC after 300 samples. The power of the additive noise was 0.01, representing a signal to noise ratio of approximately 21dB. It is clear that for both structures $P_{(2,0)}(1)$ lies within the shaded region and $P_{(2,0)}(-1)$ lies within its complement which indicates that both MLP and VSC are correctly reconstructing the input sequence, though they employ radi-

cally different solutions to do so. We further remark that the VSC requires many fewer samples to learn the particular task.



(a) MLP.



(b) VSC.

Figure 8. Decision regions for nonminimum phase channel.

We recall from Section 2, the situation where the channel is minimum phase, so that there exists an integer m such that $P_{(m,0)}(1)$ and $P_{(m,0)}(-1)$ are linearly separable, but the signal to noise ratio is poor. In Figure 9 we see the respective decision regions formed by a 2-9-5-1 perceptron and (2,5) VSC equaliser, equalising the channel with response $1.0 + 0.5z^{-1}$, following a training period of 300 samples duration. The power of the additive noise, n^2 , was 0.2, representing a signal to noise ratio of approximately 8dB. The solid curves in Figure 9 show the position of the decision boundary formed by the optimal equaliser discussed in a previous section. It can be seen that the decision region formed by the perceptron coincides in the region of interest with that of

the optimal equaliser, which suggests that the perceptron is utilising the available information with something approaching maximum efficiency. Likewise the VSC decision region shows a close correspondence with the optimal equaliser. This suggests that the performance of the nonlinear structures should compare favourably with the performance of their linear counterpart, and this is verified by the following simulations.

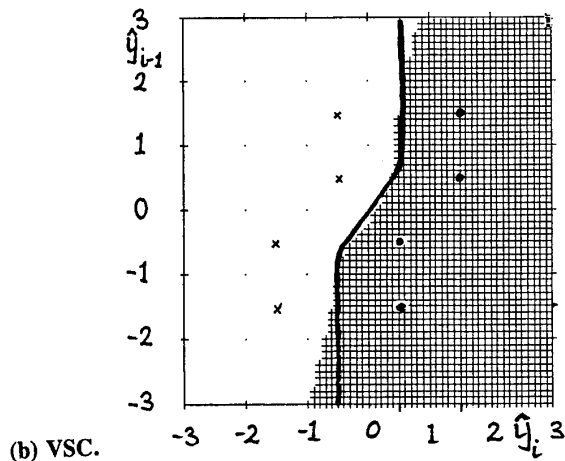
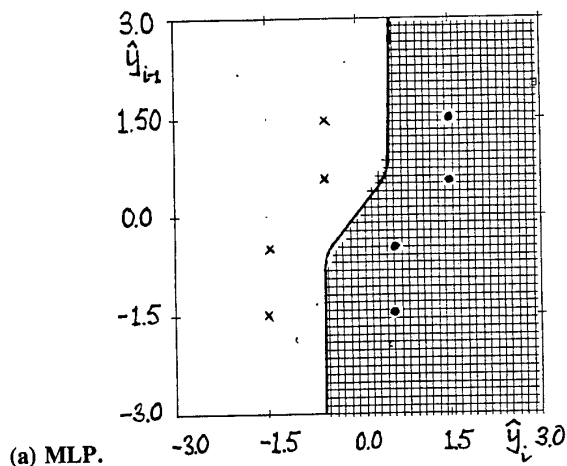


Figure 9. Decision regions for minimum phase channel.

We now turn our attention to a more realistic situation which is more representative of the kind of channel which may be encountered in practice. The simulations which we now present demonstrate clearly the superior performance enjoyed by the perceptron equaliser in comparison with the LTE. Figure 10 compares the respective BERs achieved by a 5-9-3-1 perceptron, a (5,3) VSC and an LTE of order 5, when the equalisers are set the task of equalising the channel with response function $0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$. All three structures operate with delay $d=1$, the sets $P_{(5,1)}(-)$ being linearly separable.

The simulations clearly show that the nonlinear classifiers enjoy a superior performance, in terms of the

BER achieved, to that of the LTE over the range of signal-to-noise ratios considered.

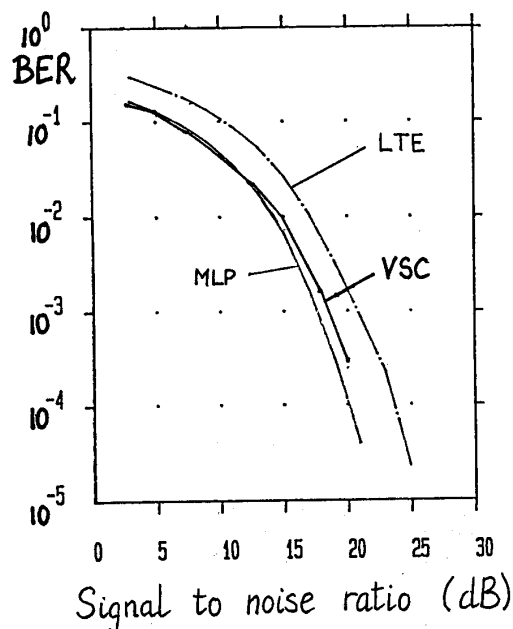


Figure 10. Comparison of bit-error rates achieved by MLP, VSC and LTE.

5. CONCLUSIONS

By incorporating a degree of nonlinearity in the design of an equaliser it is possible to produce a structure which can achieve a BER superior to that offered by linear structures, when equalising bipolar signals passed through linear channels, especially when signal to noise conditions are poor. It may be that these structures also have some value in overcoming problems associated with the equalisation of nonminimum phase channels. The multilayer perceptron has been cited as one architecture capable of realising this improvement, and the computer simulations presented in this paper suggest that this is justified. However at present there is little known concerning the theoretical analysis of such networks and the high dimensionality and degree of nonlinearity make this a difficult proposition. An alternative structure which uses polynomial nonlinearities has also been proposed as an adaptive equaliser, and these may prove to be more theoretically tractable than the perceptron. In summary, we feel justified in claiming that nonlinear structures in general can offer advantages over their linear counterparts in the design of adaptive equalisers and are certainly worthy of further investigation.

ACKNOWLEDGEMENTS

This research was supported by the British Science and Engineering Research Council under contract GR/E10357, the National Science Council and Telecommunications Labs of the Republic of China.

- [1] Qureshi, S. U. H., "Adaptive Equalisation", Proc. IEEE, vol.73, no.9, pp.1349-1387, Sept. 1985.
- [2] Proakis, J. G., "Digital Communications", McGraw Hill, 1983.
- [3] Minsky, M., and Papert, S., "Perceptrons: An Introduction to Computational Geometry", MIT Press, 1969.
- [4] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning Internal Representations by Error Propagation", in Rumelhart, D. E. and McClelland, J.L. (eds), "Parallel Distributed Processing: Explorations in the Microstructure of Cognition", MIT Press, 1986.