DYNAMICAL FUNCTIONAL ARTIFICIAL NEURAL NETWORK: USE OF EFFICIENT PIECEWISE LINEAR FUNCTIONS

J. L. FIGUEROA and J. E. COUSSEAU

CONICET -Departamento de Ingeniería Eléctrica y de Computadoras. Universidad Nacional del Sur
Avda. Alem 1253, 8000 Bahía Blanca, Argentina. jcousseau@uns.edu.ar

Abstract— **A nonlinear adaptive time series predictor has been developed using a new type of piecewise linear (PWL) network for its underlying model structure. The PWL Network is a D-FANN (Dynamical Functional Artificial Neural Network) the activation functions of which are piecewise linear. The new realization is presented with the associated training algorithm. Properties and characteristics are discussed. This network has been successfully used to model and predict an important class of highly dynamic and non-stationary signals, namely speech signals.**

Keywords— **Adaptive signal processing, nonlinear prediction, time series prediction.**

I. INTRODUCTION

The prediction of a time series can be closely related to the modeling of the underlying mechanism responsible for its generation. Many of the real physical signals encountered in practice have two characteristics: nonlinearity and non-stationarity. Consider, for example, the case of speech signals. It is known that the use of prediction plays a key role in the modeling and coding of speech signals (Shuzo and Nakata, 1985). The production of a speech signal is the result of a dynamic process that may be both non-stationary and nonlinear. To deal with the non-stationary nature of speech signals, the customary practice is to invoke the use of adaptive filtering. However, the nonlinear modeling of the speech production process is of recent vintage (Eltoft and de Figueiredo, 2000) and continues to be a research topic of active interest. As a sample of the interest in the topic, the results of a special competition looking for improved performance in times series prediction were presented recently in Lendasse *et al*. (2007).

In recent years, several structures have been developed for identification of nonlinear systems, and modeling and prediction of time series. Among these, the conventional (Schetzen, 1981) and Generalized Fock Space (de Figueiredo and Dwyer, 1980; de Figueiredo, 1983; Zyla and de Figueiredo, 1993) models of the Volterra series, the multilayer perceptron (Knecht, 1994), and the radial basis functions network (Chen *et al.*, 1991) are some of the more evident. Other works in applying neural networks for time series also include Werbos (1988), Weigend *et*

al. (1990) and de Figueiredo (1993). In Haykin and Li (1995) a pipeline recurrent neural network formed by a cascade of recurrent neural network was proposed.

A class of neural networks especially relevant to the developments in this paper is that of Dynamical Functional Artificial Neural Networks (D-FANNs). D-FANNs are artificial neural networks in which the synapses are represented by linear filters rather than memoryless links with prescribed gains or weights. For continuous-time systems, D-FANN structures without being so called, were introduced by Zyla and de Figueiredo (1993). They were reiterated as neural networks by Newcomb and de Figueiredo (1996). In 1998, generic D-FANNs both for the continuous-time and discrete-time cases were proposed and investigated by de Figueiredo (1998b). In 2000, Eltoft and de Figueiredo (2000) proposed a D-FANN for nonlinear time series prediction in which the synapses of the first layer are implemented by a filter bank built up of discrete cosine transform basis functions (DCT) and the activation functions of the first layer are smooth nonlinear functions (such as *tanh*(*x*)).

This work is addressed to the class of D-FANNs in which the synapses of the first layer are FIR filters and the activation functions are piecewise linear functions. It is presented here a study of an improved version of that class, a Piecewise Linear (PWL) D-FANN, that contemplates a recently proposed basis for the PWL representation. In addition, the PWL description in the present work includes saturation when the input signal exceeds the considered domain and allows to a more selective effect of the parameters on particular regions. In this way, we obtain good convergence properties and low complexity in terms of the number of parameters involved in the realization. Also, an associated learning algorithm is presented that leads to robust results in terms of convergence speed. Preliminary results on this subject by the authors were presented in Figueroa *et al*. (2002).

The paper is organized in the following manner. In Section 2 some concepts on time series prediction are briefly reviewed. The PWL-DFANN structure is presented and its properties are introduced in Section 3. In addition, an algorithm for training the network is discussed in Section 4. In Section 5, we present examples to illustrate the characteristics and performance of the proposed realization in terms of convergence and com-

plexity in the prediction of speech signals. The paper concludes with some final remarks in Section 6.

II. TIME SERIES PREDICTION

Following the developments by Eltoft and de Figueiredo (2000), we note that forward prediction of a time series $\{x(k)\}$, where k is the time index, can be defined as follows: Given a finite sequence of samples of a discrete time series $x(k)$, *i.e.*, $x(k)$, $x(k\text{-}1)$, … , $x(k\text{-}M)$, find the continuation $x(k+1)$, $x(k + 2)$, … . This involves finding a scalar M and a function *f*, such that $x(k + 1)$ can be estimated by

$$\hat{x}(k+1) = -f\big(x(k),\cdots,x(k-M)\big). \tag{1}$$

This is equivalent to model the time series as

$$\hat{x}(k+1) + f\big(x(k),\cdots,x(k-M)\big) = -n(k+1), \tag{2}$$

where $n(k+1)$ is a white noise process. If the statistics of the time series $x(k)$ is non-Gaussian or the time series is the result of some nonlinear operation, the function $f(.)$ is nonlinear. Equation (1) defines a *generic nonlinear AR model* and can be expressed concisely in the form

$$\hat{x}(k+1) = -f(\mathbf{x}^k), \tag{3}$$

where $\mathbf{x}^k = \big[x(k), x(k-1),\cdots,x(k-M)\big]^T$.

III. PWL-DFANN STRUCTURE

In this section, the basic structure for the PWLDFANN is described, and its general properties are discussed.

The PWL-DFANN for time series prediction is defined as a parallel connection of the *L* PWL neurons (PWLN) as illustrated in Fig. 1. Each PWLN performs the mathematical operations shown in Fig. 2, where the activation function is a piecewise linear (PWL) function on the interval comprising its domain. Specifically, the activation function for neuron q, is represented as

$$y_q(k) = g(v_q(k)) = \mathbf{c}_q^T \Lambda(v_q(k)), \tag{4}$$

where

$$\mathbf{c}_q^T = \begin{bmatrix} c_0^q & c_1^q & \cdots & c_\sigma^q \end{bmatrix}^T, \tag{5}$$

$$\Lambda(v_q) = \begin{bmatrix} 1 \\ \Lambda_1(v_q(k)) \\ \vdots \\ \Lambda_\sigma(v_q(k)) \end{bmatrix}, \tag{6}$$

with

$$\Lambda_i(v_q(k)) = \begin{cases} \tfrac{1}{2}\big(v_q(k) - \beta_i + \big|v_q(k) - \beta_i\big|\big); v_q \le \beta_\sigma \\ \tfrac{1}{2}\big(\beta_\sigma - \beta_i + \big|\beta_\sigma - \beta_i\big|\big); v_q > \beta_\sigma \end{cases} \tag{7}$$

Figure 3 illustrates the contribution of each component in the definition of a general PWL function. Note that the parameters β_i ($i = 1, ..., \sigma$) define the partition of the PWL function (Figueroa *et al.*, 2004).

The linear combination of the inputs of PWLN q represents an *M*-order FIR structure, that can be written as

$$v_q(k) = \sum_{j=0}^{M} h_j^q x(k - j) = \mathbf{h}_q^T \mathbf{x}^k, \tag{8}$$

where $\mathbf{h}_q = \big[h_0^q\, h_1^q \cdots h_M^q\big]^T$. Indeed, using the description for each PWLN given by (8) and (4), the prediction (3) can be written as

$$\hat{x}(k+1) = -\sum_{q=1}^{L} \mathbf{c}_q^T \Lambda\big(\mathbf{h}_q^T \mathbf{x}^k\big), \tag{9}$$
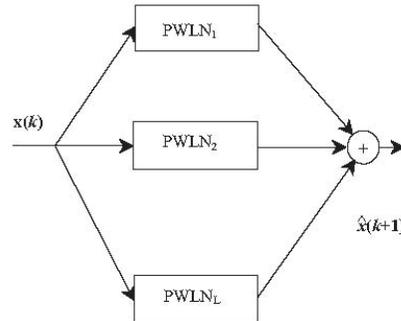

Figure 1: PWL-DFANN realization for time series prediction.
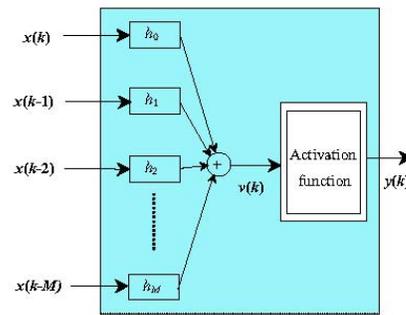

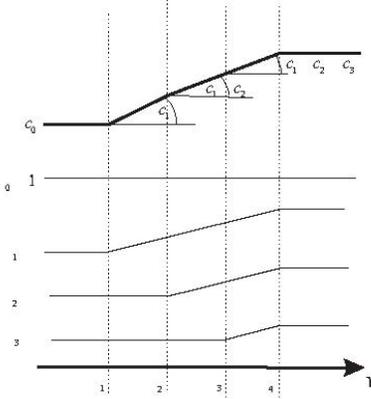Figure 2: Basic structure of the PWL neuron.

Figure 3: Effect of each parameter on the description of PWL function.

where the parameters to be estimated are the FIR filter bank coefficients ($\mathbf{h}_q$) and the parameters of the PWL functions ($\mathbf{c}_q$).

Note that the prediction error for this structure is given by

$$e(k) = x(k+1) + \sum_{q=1}^{L} \mathbf{c}_q^T \Lambda \left(\mathbf{h}_q^T \mathbf{x}^k \right) \cdot \qquad (10)$$

Some particular aspects observations related with this structure are the following,

- The largest improvement in the definition of the PWLN if compared with any classical neuron found in the literature, is that in our case, the nonlinearity can be adjusted for each particular application. Let us study this fact in detail. From Fig. 2 it is clear that the proposed PWLN can be thought as a linear FIR filter cascaded with a nonlinear gain. This realization characterizes the D-FANN nature of the proposed neural network. It was shown in de Figueiredo (1998a) that one can obtain a similar D-FANN as a best approximation to the Wiener model. For a recent such interpretation, see Figueroa and Cousseau (2001).

- A complete analysis of the particular network of a single neuron, with a detailed analysis of its properties, can be found in Figueroa *et al.* (2004).

- The use of PWL to represent the activation function allows the representation of any well-behaved continuous nonlinear function. In fact, a piecewise linear function is an approximate representation of a nonlinear function. It substitutes the global nonlinear function by a series of linear sub-functions which are defined in properly partitioned sub-regions of the original nonlinear function domain. Traditionally, a general expression for this representation is given by (Lin and Unbehauen, 1990) $f(\mathbf{x}) = a + \mathbf{b}^T \mathbf{x} + \sum_{j=1}^{\sigma} c_j \left| \alpha_j^T \mathbf{x} - \beta_j \right|$, where $\mathbf{b}$ and α_j ($j=1, \ldots, \sigma$) are M-dimensional weight vectors, a and β_j ($j=1, \ldots, \sigma$) are scalar weights. Geometrically, this function divides the input space in regions, and for each region, a linear affine model represents the system. This representation has found extensive use in the study of nonlinear circuits and systems, but can only represent nonlinearities with domain in R1 (Julián *et al.*, 1999).

Although the use of PWL descriptions is not new (Fujisawa and Kuh, 1972; Girosi *et al.*, 1994), we choose for the PWL-DFANN a recently proposed representation that allows a very compact parameterization of the realization. We assume the PWL description as defined in Julián *et al.* (1999). This description is based on a *simplicial partition* ($v = \beta_j$, with the β_j values dividing the domain in equal partitions). As a result, it is easy to verify that fixing any set of adjustable parameters ($\mathbf{h}_i$ or $\mathbf{c}_i$), the approximation error is linear in the other set. These facts lead to a very low complexity realization and a simple associ-

ated training algorithm, as will be discussed in the next section.

- Compared with preliminary studies (Figueroa *et al.*, 2002), the PWL description in the present work includes saturation when the input signal exceeds the considered domain and allows a more selective effect of the parameters on particular regions (*i.e.*, each entry in the vector $\mathbf{c}_i$ involves only the function values on a portion of the domain). These two characteristics lead to improved convergence properties of the training algorithms.

- The PWL requires the selection of the partition parameters β_i for $i = 1, \ldots, \sigma$. The interval $[\beta_1, \beta_\sigma]$ must contain the range of the signal $v(k)$, keeping in mind that the parameters $\mathbf{h}$ are time varying. This is solved by choosing the interval $[\beta_1, \beta_\sigma]$ wide enough with respect to the variation of the linear filter parameters and the input-signal range. After choosing $[\beta_1, \beta_\sigma]$, it remains to determine the interior points. Despite that some specific application could demand for a particular (irregular) density of the interior points (Hagenblad, 1999), the common sense is to use an uniform distribution for these points. This will be the usual choice in the application examples illustrated in next sections.

- Selection of the number of neurons L, is equivalent to the selection of the number of neurons in any traditional neural network. This is in general a non trivial task because there are no generally acceptable theories for the subject, and the solutions available in the literature are valid only for special cases. Usually, it is recommended to start with a small L, and if the fitting obtained is not fair, the number of neurons is increased. In general, a small L leads to an insufficient number of parameters to characterize the model and therefore a poor performance. On the other hand, a large L will give a good training and a bad generalization due to over fitting.

- Another improvement of the present formulation over the preliminary studies carried out by the authors is related to the number of parameters. In Figueroa *et al.* (2002), the realization is formed by the linear combination of the neurons. However, in order to reduce the complexity of the realization, the linear coefficients considered there (called w_i) can be modeled using the parameters of the PWL function ($\mathbf{c}_i$), without loss in the approximation capabilities.

- Note that within the space of PWL structures considered, the current method allows the best nonlinear approximation (in the least squares sense) of the desired predictor (de Figueiredo, 2000).

IV. ALGORITHM FOR PWLN TRAINING

In this section it is presented an adaptive algorithm to adjust the parameters to the time series data. To this purpose, the objective is to minimize the mean squared

189

prediction error given by

$$E\left[e^2(k)\right]= E\left[\left(x(k+1)+\sum_{q=1}^{L}\mathbf{c}_q^T\Lambda\left(\mathbf{h}_q^T\mathbf{x}^k\right)\right)^2\right].\qquad(11)$$

adjusting the parameters $\mathbf{h}_q$ and $\mathbf{c}_q$, for $q = 1, \ldots, L$. We consider the following particular updating scheme for each set of parameters. Since the error signal is linear in the set of parameters $\mathbf{c}_q$, we proposed a Recursive Least Squares (RLS) algorithm for their estimation. On the other hand, since the proposed realization is conceptually related to the structure introduced in Eltoft and de Figueiredo (2000), where a DCT is used for the (fixed) linear part, and in order to maintain a low complexity realization, an stochastic gradient algorithm is proposed for the FIR coefficients $\mathbf{h}_q$.

The design of an RLS algorithm (Haykin, 1996) for $\mathbf{c}_q$ parameters is straightforward. To that purpose it is convenient to compile these parameters in a single matrix, $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \ldots\ \mathbf{c}_L]$, and also to define the following vector,

$$\Lambda = \left[\Lambda(\mathbf{h}_1^T\mathbf{x}^k)\quad \Lambda(\mathbf{h}_2^T\mathbf{x}^k)\quad \cdots \quad \Lambda(\mathbf{h}_L^T\mathbf{x}^k)\right]^T.\qquad(12)$$

On the other hand, the stochastic gradient algorithm used to update $\mathbf{h}_q$ is described by

$$\mathbf{h}_q(k+1) = \mathbf{h}_q(k) - 2\mu e(k)\frac{\partial e(k)}{\partial \mathbf{h}_q(k)},\qquad(13)$$

where μ is the step-size. In our description, the gradient can be computed as

$$\frac{\partial e(k)}{\partial \mathbf{h}_q(k)} = \left(\mathbf{c}_q^T\frac{\partial \Lambda(v_q)}{\partial v_q}\right)\mathbf{x}^k,\qquad(14)$$

where, if $[.]_j$ represents the j-th vector component, and

$$\left[\frac{\partial \Lambda(v_q)}{\partial v_q}\right]_1 = 0,\qquad(15)$$

$$\left[\frac{\partial \Lambda(v_q)}{\partial v_q}\right]_{j+1} = \begin{cases}\frac{1}{2}\left(1+sign\left(v_q(k)-\beta_j\right)\right); & v_q \le \beta_\sigma \\ 0 & ; v_q > \beta_\sigma\end{cases}.\qquad(16)$$

Regarding a suitable definition of PWL gradient at the partition edges, the gradient at the partitions boundaries is defined as zero to avoid any numerical inconsistency. In Eq. 16 it is used that $sign(0)=-1$.

The complete learning algorithm for the PWL-DFANN realization is presented in Table 1.

In order to guarantee convergence of the coefficients in the mean, the step-size of the LMS algorithm must be chosen as a small positive number that satisfy (Figueroa *et al.*, 2004).

$$0 < \mu < \min_q\left(\frac{1}{\rho\sum_{i=2}^{\sigma+1}\left(c_i^q\right)^2}\right),\qquad(17)$$

where ρ represents the maximum eigenvalue of matrix $E[(\mathbf{x}^k)^T\mathbf{x}^k]$. Note that this expression is a simple bound for μ since ρ depends on the data and, in general, it is easy to obtain a reasonable upper bound for $\sum_{i=2}^{\sigma+1}\left(c_i^q\right)^2$.

Table 1: Learning algorithm for the PWL-DFANN realization.

Parameters: L = number of FIR filters
M = order of FIR filters
σ = number of partitions
μ = step-size related to h coefficients
λ = forgetting factor related to c coefficients
δ = initialization of $\mathbf{P}(0)$
Data: x
Initialization:
$\quad\mathbf{h}_q(0)$ = DCT coefficients
$\quad\mathbf{c}_q(0) = [-1\ 1\ 0\ \ldots\ 0]^T\cdot$
$\quad\mathbf{P}(0) = \delta^{-1}\mathbf{I}$
For each $k = 1, 2, \ldots$
$\quad$ For $q = 1, \ldots, L.\cdots$
$\qquad v_q(k) = \mathbf{h}_q(k)^T\mathbf{x}^k$
$\quad\hat{x}(k+1) = \mathbf{C}^T\Lambda$
$\quad e(k) = x(k+1) - \hat{x}(k+1)$
$\quad$ For $q = 1, \ldots, L$
$\qquad \mathbf{h}_q(k+1) = \mathbf{h}_q(k) - 2\mu e(k)\dfrac{\partial e(k)}{\partial \mathbf{h}_q(k)}$
$\quad \mathbf{g}(k) = \dfrac{\lambda^{-1}\mathbf{P}(k-1)\Lambda}{1+\lambda^{-1}\Lambda^T\mathbf{P}(k-1)\Lambda}$
$\quad \mathbf{C}(k+1) = \mathbf{C}(k) + \mathbf{g}(k)e(k)$
$\quad \mathbf{P}(k) = \lambda^{-1}\mathbf{P}(k-1) - \lambda^{-1}\mathbf{g}(k)\Lambda^T\mathbf{P}(k-1)$

An important aspect for this updating algorithm, and for any nonlinear adaptive filter, is the selection of the initial condition for the parameters. In particular, considering the same parameters for all FIR filters $\mathbf{h}_q$, this will move to a ill-conditioned problem. To avoid this, a good choice for the initial condition of these parameters is the DCT basis proposed in Eltoft and de Figueiredo (2000). In addition to avoid the ill-conditioned problem, this selection can take profit of the orthogonality properties of these filters.

In the next section, the proposed PWL-DFANN scheme is applied in the context of speech prediction and illustrative practical results are discussed. .

V. SIMULATION EXAMPLES

Speech signals are an important class of signals that, on a short time period (5-100 ms), have statistical properties that are slowly varying. On the other hand, over longer periods of time (of the order of 1/5 s) they are highly dynamical and non-stationary. To illustrate the performance of the proposed realization, we present in this section examples of the use of PWL-DFANN realization for time series prediction speech signals.

A. Speech signals
In this section we will apply the PWL-DFANN realization for time series prediction to predict the next sample of the speech signal depicted in Fig. 4. The recorded time series is made up of 10000 samples, sampled at 8 kHz. This is the same sample signal studied in Haykin and Li (1995) and in Eltoft and de Figueiredo (2000), and is selected for specific comparison purposes.

The input signal segment was projected on to 12

filters of 12-order (Eltoft and de Figueiredo, 2000)) ($L = 12$, $M = 12$). The activation function was approximated using a partition of 10 sectors, to allow a smooth approximation ($\sigma = 10$). Looking for fast convergence, the step-size of the LMS algorithm is set equal to $\mu = 0.02$, the forgetting factor of the RLS algorithm is set equal to $\lambda = 0.998$ (similar to the value used in Eltoft and de Figueiredo, 2000), and the initial correlation matrix coefficient is fixed as $\delta = 50$. The squared prediction error is depicted in Fig. 5. For quantitative comparison with other prediction algorithms, we used the following performance index introduced by Haykin and Li (1995),

$$R_p = 10 \log_{10} \left(\frac{\delta_s^2}{\delta_p^2} \right)^k , \qquad (18)$$

where δ_s^2 is the mean square value of the incoming signal, and δ_p^2 is the corresponding value of the prediction error.

Comparison results using previous index are summarized in Table 2, where other values found in the literature are also included for this example. Nominally, a linear adaptive predictor (LAP), a *pipeline recurrent neural network* (PRNN; Haykin and Li, 1995), a multilayer perceptron (MP) and the DFANN (Eltoft and de Figueiredo, 2000).
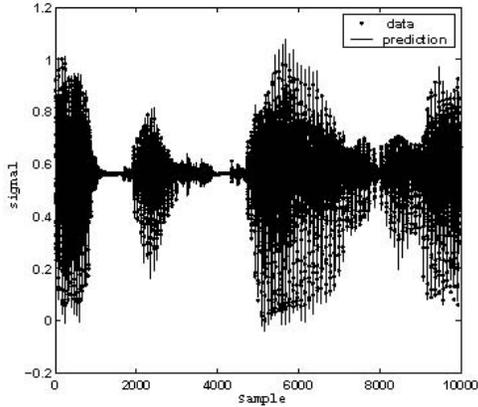

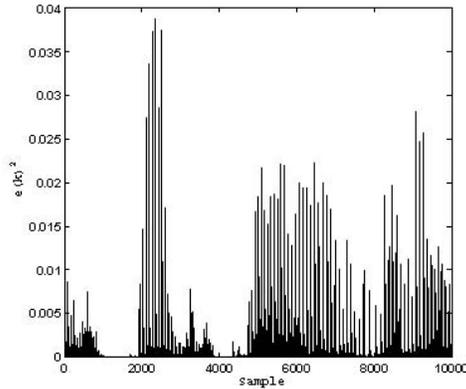Figure 4: Speech signal and prediction obtained using PWL-DFANN.

Figure 5: Squared prediction error using PWL-DFANN realization.

As can be concluded from this, and other exhaustive computer simulations, the index performance obtained for the PWL-DFANN is highly competitive if compared with the other approaches. Table 2 includes also the number of parameters to be updated in each implementation[1]. An important aspect to consider is the benefit of the adaptation of parameters **h**. Note that the DFANN (Eltoft and de Figueiredo, 2000) uses a fixed set of parameters. When we consider these parameters fixed at the initial condition, the performance obtained is given by $R_p = 27.81$. This value is slightly lower than the performance obtained when these parameters are adjusted. As a consequence, maintain fixed **h** parameters is an interesting alternative, mostly because the set of parameters to be adjusted is reduced to 132. As mentioned above, the initial values of h are set as the DCT basis (Eltoft and de Figueiredo, 2000).

Figure 6 illustrates the dependence of Rp with the number of neurons L. It can be observed with this figure that with an increase of L leads to an improvement in the index performance Rp. However, use of large L (higher than 12 for this example) leads in general to produce over-fitting. Figure 7 illustrates the dependence of performance index Rp with the step-size μ. From this figure it is clear that, for reasonable values of the step size μ, the adaptation of h allows a small improvement of the performance index.

Finally, a study of the dependence on the estimation of **C** of the RLS forgetting factor λ is illustrated in Fig. 8. In this plot the values of R_p are depicted for several values of λ. As can be concluded, the selection of λ is not critical to obtain a good performance using the PWL-DFANN.

B. Handel's Hallelujah Chorus

In this example a set of 72000 samples, extracted from the Handel's Hallelujah Chorus, is used for quantitative comparison. The prediction algorithms used in the comparison are: a linear adaptive predictor (LAP), the *pipeline recurrent neural network* (PRNN; Haykin and Li, 1995), the D-FANN (Eltoft and de Figueiredo, 2000) and the proposed PWLDFANN. Comparison results, using previous performance index, are summarized in Table 3.

Table 2: Comparative results for different non linear predictors in Example 1.

Technique	R_p	Parameters
PWL-DFANN	27.8945	276
DCT-DFANN	27.55	36 [†]
PWL-DFANN[‡]	27.2869	288
PRNN	25.14	52
LAP	22.01	12
MP	≈ 18	169[*]

[†] Plus 144 fixed coefficients taken as DCT.; [‡] Figueroa *et al.* (2002); [*] MP with 12 neurons on the hidden layer

[1] The results used for comparison purposes are taken from the literatura (Eltoft and de Figueiredo, 2000; Figueroa *et al.*, 2002; Haykin and Li, 1995).
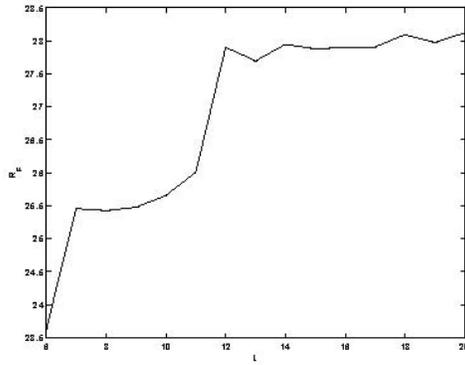
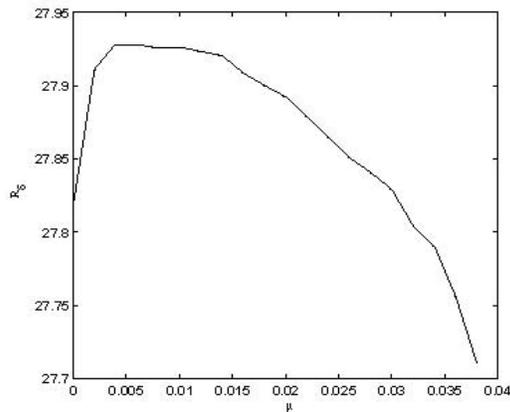Figure 6: Dependence of R_p as function of the number of neurons L.

Figure 7: Dependence of R_p as function of the step-size μ.

Table 3: Comparative results for different non linear predictors in Example 2.

Technique	R_p
PWL-DFANN	18.0771
DCT-DFANN	17.0065
PRNN	16.0899
LAP	15.0008

From this data the values of the performance index obtained for the PWL-DFANN shows to be high-ly competitive if compared with the other approaches.
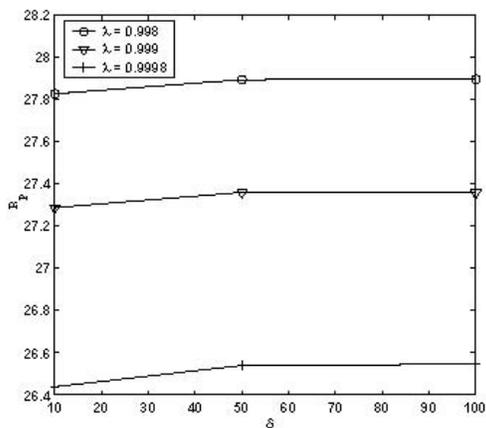

Figure 8: Dependence of R_p as function of the forgetting factor λ for the RLS-algorithm.

VI. CONCLUSIONS

In this paper, we have used a PWL-DFANN to build a nonlinear, adaptive time series predictor. The compact PWL description utilized for the PWL-DFANN includes saturation when the input signal exceeds the considered domain. This particular definition allows a selective effect of the parameters on particular regions. Training is a combination of a stochastic gradient algorithm and a RLS algorithm, leading to robust experimental results. The resulting realization is simple and the performance shows that this structure can be successfully used to perform prediction of highly non-stationary and dynamic speech signals.

REFERENCES

Chen, S., C.F.N. Cowan and P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks", *IEEE Trans. Neural Networks*, **2**, 302-309 (1991).

de Figueiredo, R.J.P. and T.A. Dwyer, "A best approximation framework and implementation for simulation of large-scale nonlinear systems," *IEEE Trans. Circuits Syst.,* **CAS-27**, 1005-1014 (1980).

de Figueiredo, R.J.P., "A generalized Fock space framework for nonlinear system and signal analysis", *IEEE Trans. Circuits Syst.*, **CAS-30**, 637-647 (1983).

de Figueiredo, R.J.P., "A neural network model for nonlinear predictive coding in Fock space," *IEEE Int. Symp. Circuits and Systems*, Chicago, IL, 2164-2167 (1993).

de Figueiredo, R.J.P., "Optimal interpolating and smoothing functional artificial neural networks (FANNs) based on a generalized Fock space framework," *Circuits, Systems, and Signal Processing*, **17**, 271-87 (1998a).

de Figueiredo, R.J.P., "Beyond Volterra and Wiener: Some new results and open problems in nonlinear signal and image processing," *Midwest Symposium on Circuits and Systems*, , 124-127 (1998b).

de Figueiredo, R.J.P., "A reproducing kernel Hilbert space (RKHS) approach to the optimal modeling, identification, and design of nonlinear adaptive systems," *IEEE Symp. on Adaptive Systems for Signal Processing, Comm. and Control*, 42-47 (2000).

Eltoft, T. and R.J.P. de Figueiredo, "A DCT-Based DFANN for nonlinear adaptive time series prediction," *IEEE Trans. Circuits Syst., Part II*, **47**, 1131-1134 (2000).

Figueroa, J.L. and J.E. Cousseau, "A Wiener CPWL Adaptive Filter," *2nd. IEEE South-American Workshop on Circuits and Systems*, SAWCAS'2001, Buenos Aires, Argentina, 2001.

Figueroa, J.L., J.E. Cousseau and R.J.P. de Figueiredo, "A Piecewise Linear Dynamical Functional Artificial Neural Network (PWL-DFANN) for Nonlinear Adaptive Time Series Prediction," *IEEE Int. Symp. Circuits and Systems*, I29-I32 (2002).

Figueroa, J.L., J.E. Cousseau and R.J.P. de Figueiredo, "A low complexity simplicial canonical piece-wise linear adaptive filter," *Circ., Syst. and Signal Proc. Journal*, **23**, 365-386 (2004).

Fujisawa, T. and E.S. Kuh, "Piecewise-linear theory of nonlinear networks," *SIAM J. Appl. Math.*, **22**, 307-328 (1972).

Girosi, F., M. Jones and T. Poggio, "Regularization theory and neural networks architecture," *Neural Computation*, **7**, 219-269 (1995).

Hagenblad, A., "Aspects of the identification of Wiener models," *M.Sc thesis dissertation*, Linkoping University, Sweden (1999).

Haykin, S., *Adaptive Filter Theory*, Prentice Hall, N.J. (1996).

Haykin, S. and L. Li, "Nonlinear adaptive prediction of nonstacionary signals," *IEEE Trans. Signal Processing*, **43**, 526-535 (1995).

Julián, P., A. Desages and O. Agamennoni, "High Level Canonical Piecewise Linear Representation Using a Simplicial Partition," *IEEE Trans. Circuits Syst., Part I*, **46**, 463-480 (1999).

Knecht, K.G., "Nonlinear noise filtering and beamforming using the perceptron and its Volterra approximation," *IEEE Trans. Acoust. Speech Signal Processing*, **2**, 55-62 (1994).

Lendasse, A., E. Oja, O. Simula, and M. Verleysen, "Times series prediction competition: The CATS benchmarck," *Neurocomputing*, **70**, 2325-2329 (2007).

Lin, J.N. and R. Unbehauen, "Adaptive nonlinear digital filter with canonical piecewise linear structure," *IEEE Trans. Circuits Syst.*, **CAS-37**, 347-353 (1990).

Newcomb, R.W. and R.J.P. de Figueiredo, "A multi-input multi-output functional artificial neural network," *J. of Intelligent and Fuzzy systems*, **4**, 207-213 (1996).

Schetzen, M., "Nonlinear system modeling based on the Wiener theory," *Proc. IEEE*, **69**, 1557-1573 (1981).

Shuzo, S. and K. Nakata, *Fundamentals of Speech Signal Processing*, Academic Press, New York (1985)

Weigend, A., B. Huberman and D. Rumelhart, "Predicting the future: A connectionist approach," *Int. J. Neural Syst.*, **7**, 403-430 (1990).

Werbos, P., "Generalization of backpropagation with application to a recurrent gas market model," *Neural Networks*, **1**, 339-365 (1988).

Zyla, L.V. and R.J.P. de Figueiredo, "Nonlinear system identification based on a Fock space framework," *SIAM J. Contr. Optimiz.*, **31**, 931-939 (1993)