

# BACKWARD ADAPTIVE RBF-BASED HYBRID PREDICTORS FOR CELP-TYPE CODERS AT MEDIUM BIT-RATES

C. Peláez-Moreno and F. Díaz-de-María

Dpto. Tecnologías de las Comunicaciones, Universidad Carlos III de Madrid  
C/ Butarque, 15; 28911 Leganés (Madrid), Spain  
carmen@ and fdiaz@tsc.uc3m.es

## ABSTRACT

Nonlinear prediction is a natural way to increase the quality of speech coders. Several approaches have been recently proposed in this direction ([1,2,3,4] are some examples) and most of them use neural networks as predictors. Nevertheless, the computational cost due to the network training is very high, since it usually involves a gradient descent-based nonlinear optimization process. In this paper we propose some improvements of our previous work reported in [3], all of them aiming at reducing the computational cost. Our predictor can be used in CELP-type coders and provides a 0.6 dB increase of the SEGSNR with respect to conventional CELP coders.

Keywords: RBF Network, Nonlinear Prediction, Backward Adaptive Prediction, Speech Coding.

## 1. INTRODUCTION

Toll quality speech coders at 8 kb/s were already feasible a few years ago. Nowadays, one of the main objectives of the speech coding research community is to achieve the same quality at 4 kb/s. Perhaps this achievement is not so far away, and it even may be attained by means of some refinements in present linear predictive coders. However, it seems evident that this type of coders are already close to reach their technological limits, or, in other words, substantial advances in this subject will require a somewhat different coding paradigm.

Nonlinear prediction seems a natural way to improve the performance of present speech coders, since there are clear evidences of nonlinear phenomena in the human speech production mechanism. However, the application of nonlinear prediction has to overcome a number of problems before being a practical alternative to linear prediction; in our opinion, the high computational cost and the lack of a parsimonious representation for the predictor parameters are the main difficulties.

During the last few years, there have been several contributions to overcome the above mentioned difficulties ([1,2,3,4] are relevant examples). The first of these works [1] proposed a vector-quantized neural network-based predictor for low bit-rate coders; vector quantization provides an efficient encoding of the predictor, but the computational cost of the predictor selection process is very high (the input signal is predicted using every predictor in the codebook). Backward adaptive prediction is an alternative solution, since the predictor parameters do not have to be transmitted; nevertheless, this technique can only be applied at medium bit-rates (which provide a good quality coded speech). The coders presented in [2] and [4] are nonlinear alternatives to the LD-CELP standard at 16 kb/s; these works report good performance and moderate computational cost, but the proposed predictors can not be applied to typical CELP coders (in the LD-CELP, the codebook search is much simpler than in conventional CELP coders, since there is no adaptive library and the gain associated to the fixed one is previously computed -predicted-, outside the search procedure).

In this paper we present several improvements of our previous work [3], which presented a backward adaptive hybrid predictor consisting of a linear filter and a RBF ("Radial Basis Function") network [5,6]. Here, we report a large reduction of the computational cost with respect to our previous work [3], along with other relevant developments which will be described later.

The paper is organized as follows. In section 2, the RBF network is presented as a predictor for speech coding. Section 3 describes the hybrid predictor we propose and its training method. A nonlinear predictive CELP-type coder -based on this predictor is presented in Section 4. Simulation results are discussed on Section 5. Finally, the conclusions of this work, as well as suggestions for future work, are given in Section 6.

## 2. REGULARIZED RBF NETWORKS FOR SPEECH PREDICTION

The speech prediction problem for speech coding can be stated as follows: let  $S = \{(\mathbf{x}_i, x_{i+1}) \in \mathbf{R}^p \times \mathbf{R} \mid i = 1, \dots, N\}$  be a set of data pairs to be approximated by a function  $f$ , where each pair consist of a sample to predict,  $x_{i+1}$ , and a vector of samples,  $\mathbf{x}_i$ , corresponding to previous samples of  $x_{i+1}$ , namely  $\mathbf{x}_i = [x_i, x_{i-1}, \dots, x_{i-p+1}]^T$ , with  $p$  being the prediction order.

Most of the approaches to nonlinear predictive coding use a neural network-based predictor. There are many reasons for this choice: they learn from examples, are capable of generalizing, have parallel structures, etc. Nevertheless, as previously mentioned, the computational cost of the neural network training is very high. One of the main reasons for using RBF networks is the reduced computational cost of their training compared to other types of networks.

The RBF network, shown in Figure 1, is a single-layer network which computes the formula:

$$f(x) = \sum_{j=0}^{M-1} c_j G(\|\mathbf{x} - \mathbf{t}_j\|)$$

where  $\{G(\cdot)\}$  are RBF,  $\{\mathbf{t}_j\}$  are the RBF centers,  $\{c_j\}$  are the weights of the linear combination, and  $M$  is the number of RBF used. In particular, we use Gaussian RBF

$$G(x) = \exp\left(-\frac{x^2}{\sigma^2}\right),$$

where  $\sigma$  is its variance or width.

Another reason for choosing the RBF network is that it yields a regularized solution to the prediction problem, i.e., a solution which balances, by means of the regularization parameter ( $\lambda$ ), smoothness against fitness to the data. This is important because it makes the network respond properly to quantized excitations.

The RBF training algorithms can be divided into two large groups, usually referred to as linear and nonlinear. The training is called linear when it is

performed following this two steps: first, the centers are selected through a clustering algorithm and the width by means of some heuristic procedure (which is usually based on the previous clustering); and second, the weights are obtained solving the resulting linear problem, once the centers and width are already fixed. On the other hand, the training is nonlinear when all of the parameters are jointly optimized through a computationally expensive nonlinear optimization technique (such as a gradient descent algorithm).

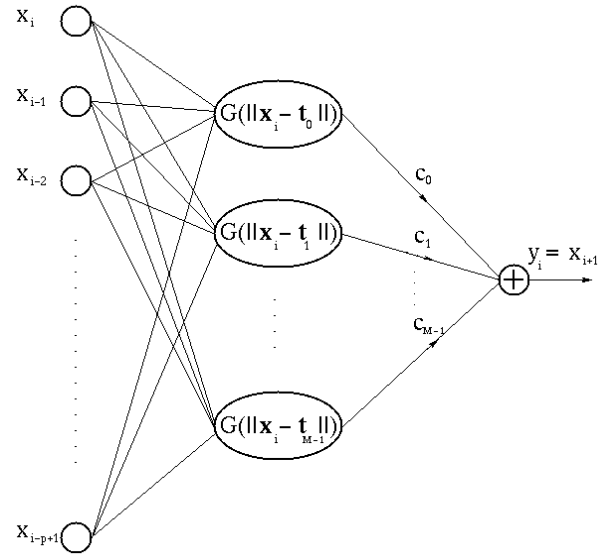


Figure 1. RBF Network

In our previous work [3] we proposed a nonlinear procedure to train the network. It leads to good results but at the expense of a high computational cost. Here we avoid this problem by showing that a linear procedure can work as good as the nonlinear one as long as a slightly bigger network is used.

In particular, the linear algorithm used in this work is as follows. First, the centers are trained using the Frequency Sensitive Competitive Learning (FSCL) algorithm [7]. Second, the width is obtained as the maximum distance among the previously computed centers. Finally, the weights are determined via a pseudoinverse (using a regularization parameter  $\lambda=10$ ).

## 3. RBF-BASED HYBRID PREDICTION

In contrast to other approaches to nonlinear prediction, we proposed to keep a linear part in the predictor for two main reasons: first, it should not be ignored that linear prediction does a good job

(though with limitations); and second, a careful design of the predictor allows us to use the efficient linear codebook search methods (by taking out the RBF network from the codebook search processes, as we will describe later).

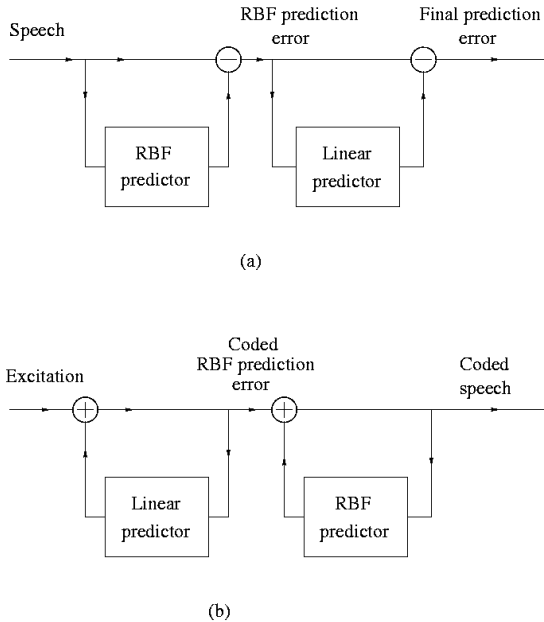


Figure 2. RBF-Based Hybrid Predictor: (a) analysis system; (b) synthesis system

The proposed hybrid predictor, shown in Figure 2(a), consists of the cascade of a RBF network and a linear predictor. A comprehensive discussion about the reasons for selecting this configuration can be found in [3]. Here we will briefly summarize the two main reasons:

- Instead of removing the linear basis and building a new global nonlinear solution, it allows to complement the linear prediction capabilities with a nonlinear contribution. Additionally, the computational cost of the nonlinear prediction is reduced.
- It can be easily applied to CELP-type coders in a suboptimal way, still providing good results. Specifically, we remove the nonlinear part of the synthesis system -see Figure 2(b)- from the excitation selection procedure: the excitation is selected to achieve the best coded RBF prediction error, i.e., the closest to the original one.

#### 4. OVERVIEW OF THE PROPOSED CODER

Figure 3 shows the block diagram of the CENP (“Code-Excited Nonlinear Predictive”) coder de-

scribed in this section, which has been built to evaluate the performance of the proposed hybrid predictor, without paying much attention to other aspects of the coder (which can be improved in several ways). For example, the perceptual weighting has been removed, since it cannot be easily included in the nonlinear coder.

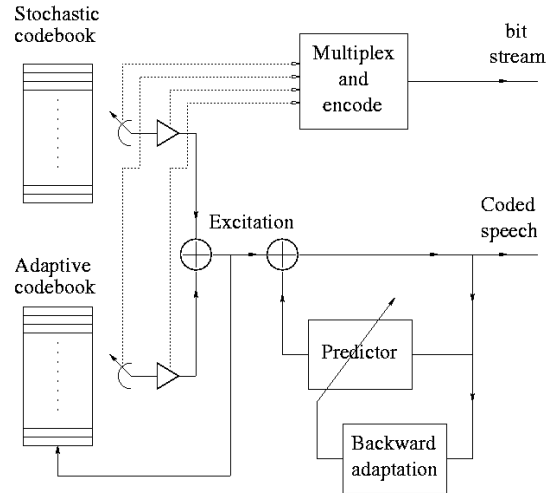


Figure 3. Block Diagram of the Code-Excited Nonlinear Predictive Coder (CENP)

As illustrated in Figure 3, the predictor adaptation is backward. Thus, the predictor parameters do not have to be transmitted. As a result, there is no bit rate increase due to the inclusion of nonlinear prediction, and quality improvements are achieved only at the cost the computational effort required to train the RBF network.

For each frame, the network centers, the width and weights are trained. Then the linear filter is adapted (using the autocorrelation method) to predict the network residual. The predictor consist of a RBF with  $M=10$  centers of dimension  $p=4$ , and  $10^{\text{th}}$  order linear predictor.

In the coder proposed in [3] the network was switched on and off depending on the result of a simplified analysis-by-synthesis procedure carried out both with and without the network. This switching has been removed in this new version, thus obtaining another reduction in the computational cost.

The long term prediction is carried out by means of an adaptive codebook. The stochastic codebook contains 1024 vectors. The analysis frame length is

160 samples (20 ms. at 8 kHz). The frame rate is 4 ms. (32 samples) and each frame is divided into two subframes of 2 ms. (16 samples). The predictor is adapted once per frame while the excitation is adapted twice per frame (at the subframe rate).

The bit allocation is shown in table I. As shown, the total number of bits per subframe is 27, leading to a bit rate of 13.5 kb/s.

PARAMETER	BITS
Stochastic Codebook Index	10
Adaptive Codebook Index	7
Stochastic Gain	5
Adaptive Gain	5
TOTAL	27

Table I. Bit Allocation

## 5. SIMULATION RESULTS

To test the performance of the suggested predictor we have compared the CENP coder with a CELP coder having the same structure and bit allocation, but using a 10<sup>th</sup> linear predictor instead of the hybrid one.

For both coders, we have computed the Segmental-SNR of 4 sentences of speech (2 of them were female and 2 male voices). Their duration and the results we have obtained are summarised in Table II.

SENTENCE	DURATION (S.)	CELP (SEGSNR)	CENP (SEGSNR)
# 1 (female)	13.8	19.75	20.53
# 2 (female)	12.5	19.62	20.49
# 3 (male)	20	13.65	13.85
# 4 (male)	17.5	19.52	20.11
Mean	15.95	18.13	18.74

Table II. Performances of the compared coders

Therefore, it can be observed that the inclusion of the non-linear predictor yields a mean improvement of 0.61 dB on the SEGSNR

## 6. CONCLUSIONS AND FURTHER WORK

In this paper we have reported several improvements of our previous work [3], where a new RBF-based hybrid predictor for CELP-type

coders was presented. Specifically, we have significantly reduced the computational cost due to the RBF network training and the simplified analysis-by-synthesis procedure used to switch on and off the network, while keeping the performance.

The achieved improvements in terms of SEGSNR with respect to conventional CELP coders (0.6 dB) are similar to those reported in [2] and [4], but our approach is applicable to more general CELP coders (including conventional search procedures).

We are currently working on the application of our predictor to the LD-CELP standard at 16 kb/s.

## 7. REFERENCES

- [1] L. Wu, M. Niranjana and F. Fallside: "Fully Vector-Quantized Neural Network-Based Code-Excited Nonlinear Predictive Speech Coding"; IEEE Trans. on Speech and Audio Processing, vol. 2, no. 4, pp. 482-489; 1994.
- [2] A. Kumar and A. Gersho: "LD-CELP Speech Coding with Nonlinear Prediction"; IEEE Signal Processing Letters; vol. 4, no. 2, pp. 89-91; 1997.
- [3] F. Diaz-de-Maria and A.R. Figueiras-Vidal: "Improving CELP Coders by Backward Adaptive Non-Linear Predictors"; International Journal of Adaptive Control and Signal Processing; vol. 2, no. 7, pp. 585-601; 1997.
- [4] N. Ma and G. Wei: "Speech Coding with Non-linear Local Prediction Model"; Proc. ICASSP-1998; vol. 2, pp. 1101-1104; Seattle, Washington, USA; 1998.
- [5] J. Moody and C. Darken, "Fast Learning in Networks of Locally-Tuned Processing Units," Neural Computation, no. 1, pp. 281-294, 1989.
- [6] S. Haykin, "Neural Networks: A Comprehensive Foundation", Englewood Cliffs, NJ: Macmillan College Publishing Company, 1994.
- [7] S. Ahalt, A. Krishnamurthy, P. Chen, D. Melton, "Competitive Learning Algorithm for Vector Quantization," Neural Networks, vol. 3, pp. 277-290; 1990.