NONLINEAR PREDICTION BASED ON SCORE FUNCTION

Jordi Solé i Casals (1), Enric Monte i Moreno (2)

(1)Signal Processing Group. University of Vic, Sagrada Família 7, 08500, Vic (Catalunya, Spain)
(2) TALP. Polytechnical University of Catalonia, Jordi Girona, 1-3, 08034, Barcelona (Catalunya, Spain)

ABSTRACT

The linear prediction coding of speech is based in the assumption that the generation model is autoregresive. In this paper we propose a structure to cope with the nonlinear effects presents in the generation of the speech signal. This structure will consist of two stages, the first one will be a classical linear prediction filter, and the second one will model the residual signal by means of two nonlinearities between a linear filter. The coefficients of this filter are computed by means of a gradient search on the score function. This is done in order to deal with the fact that the probability distribution of the residual signal still is not gaussian. This fact is taken into account when the coefficients are computed by a ML estimate. The algorithm based on the minimization of a high-order statistics criterion, uses on-line estimation of the residue statistics and is based on blind deconvolution of Wiener systems [1]. Improvements in the experimental results with speech signals emphasize on the interest of this approach.

1. INTRODUCTION

It is known that in the speech production mechanism there are present several nonlinearities [2]. These non-linealirities have been exploited for speech coding [3] [4] [5] [6] [7], in order to improve the prediction gain. In this paper we propose a new approach based in two prediction stages. The first stage (see figure 1), removes all the components that can be predicted by a linear model, and in the second stage, the signal is filtered after being compressed by a nonlinear function *g*(.) and then, goes through another nonlinear function *h*(.) (see figure 2). The idea of this second stage is to concentrate on the low energy components of the residual signal, which still have short term dependencies.

2. CLASSICAL LPC

The classical LPC methods are based on the minimization of a mean square error, defined as the difference between the input signal $x(k)$ and the predicted signal

$y(k)=[w(z)]x(k-1)$, where $w(z)$ is a L-th order causal finite impulse response filter, i.e. a filter whose entries $w_i = 0$ for $i \notin \langle 0,\ldots,L-1\rangle$. The block diagram of a linear predictor is shown in figure 1.

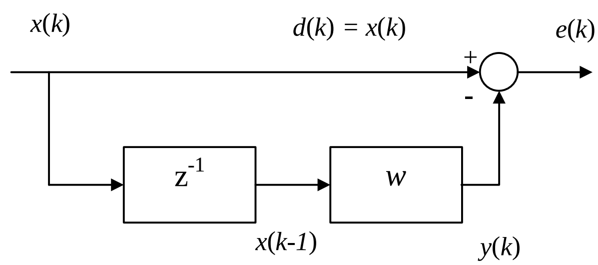

Figure 1: Block diagram linear prediction coding system.

Denoting $E[x(k)x(k-l)] = R_{xx}(l)$, the cost function reduces to:

$$J = E\left[e^2(k)\right] = R_{xx}(0) - 2\sum_{n=0}^{L-1} w_n R_{xx}(n+1) + \sum_{m=0}^{L-1}\sum_{n=0}^{L-1} w_m w_n R_{xx}(m-n) \qquad (1)$$

This estimation can be viewed as a maximum likelihood (ML) estimate in the special case of independent and identically distributed (iid) gaussian error: in fact, first consider only the prediction at time k. Taking into account the relation $y(k) = x(k) - e(k)$, and denoting

$p_E(.)$ the probability density function (pdf) of the residue $e(k)$, the log-ML estimation is :

$$ArgMax_{\mathrm{w}}\left[\sum_{i=0}^{N-1}\ln\left(p_E\left(e(k+i)\right)\right)\right] \qquad (2)$$

Assuming that the error $e(k)$ is a gaussian zero mean random variable, the Maximum Likelihood estimation is:

$$ArgMin_{\mathrm{w}}\left[\sum_{i=0}^{N-1}\left(e(k+i)\right)^2\right] \qquad (3)$$

As it is well known, in the gaussian case, asymptotically, the ML is nothing but the minimum mean square error (MMSE) estimate.

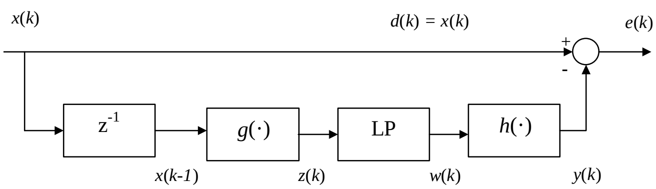

Figure 2: Block diagram of the second stage.

3. SCORE FUNCTION METHOD FOR THE LPC COEFFICIENTS

Unfortunately, if the error is not gaussian, the MMSE estimate is no longer equal to the ML estimate, as showed in [7]. In fact, from (2), one can compute the ML equation by deriving the equation with respect to the entries w_j :

$$\sum_{i=0}^{N-1}\frac{\partial}{\partial w_j}\ln\left(p_E\left(e(k+i)\right)\right)=\sum_{i=0}^{N-1}\frac{p'_E\left(e(k+i)\right)}{p_E\left(e(k+i)\right)}\frac{\partial e(k+i)}{\partial w_j}$$
$$=-\sum_{i=0}^{N-1}\psi_E\left(e(k+i)\right)x(k+i-j-1) \qquad (4)$$

where $\psi_E(.)$ denotes the derivative of $\ln p_E(.)$, the so-called score function. Consequently, asymptotically, for any error distribution, the ML estimate of w_j, $j=0,\ldots,L-1$, is equivalent to the equation set:

$$E\left[\psi_E(e(k))x(k-j-1)\right]=0, \quad j=0,\ldots,L-1. \qquad (5)$$

Basically, the score function is a nonlinear function, except in the gaussian case. In these case, $\psi_E(e)=-e$. Then, equation (5) prove that the optimal ML estimate involves higher (than 2) order statistics, except in the gaussian case.

4. SCORE FUNCTION METHOD FOR THE NONLINEAR STAGE

As can be seen in figure 2, now the error sequence is defined as:

$$\begin{aligned} e(n) &= x(n)-y(n) \\ &= x(n)-h\left(\sum_{k=0}^{L-1}f(k)g\left(x(n-k-1)\right)\right) \end{aligned} \qquad (6)$$

where $f(k)$ are the coefficients of the LP filter, and $g(.)$, $h(.)$ are nonlinear functions. We will define the auxiliary variables:

$$z(n)=g\left(x(n-1)\right);\ w(n)=\sum_{k=0}^{L-1}f(k)g\left(x(n-k-1)\right)$$

therefore, $y(n)=h\left(w(n)\right)$.

In order to compute de minimum of the cost function we need the derivative of (6) with respect to the coefficients f_k:

$$\sum_{i=0}^{N-1}\frac{\partial}{\partial f_j}\ln\left(p_E\left(e(n-i)\right)\right)=\sum_{i=0}^{N-1}\frac{p'_E e(n-i)}{p_E e(n-i)}\ \frac{\partial e(n-i)}{\partial f_j} \qquad (7)$$
$$=-\sum_{i=0}^{N-1}\Psi_E\left(e(n-i)\right)\frac{\partial h\left(w(n-i)\right)}{\partial w(n-i)}\ g\left(x(n-i-j-1)\right)$$

Consequently, asymptotically, for any error distribution, the ML estimate of f_j, $j=0,\ldots,L-1$, is equivalent to the equation set:

$$E\left[\psi_E(e(k))\frac{\partial h\left(w(k)\right)}{w(k)}g\left(x(k-j-1)\right)\right]=0, \quad j=0,\ldots,L-1. \qquad (8)$$

As we can see, comparing eq. (8) with eq.(5), we introduce high-order statistics in the equation set by means of nonlinear functions $g(\cdot)$ and $h(\cdot)$.

The update of the coefficients f is done by means of a gradient search, i.e.

$$f_k \leftarrow f_k - m \frac{\partial J(e(n))}{\partial f_k}$$

5. SELECTION OF THE NONLINEAR FUNCTIONS

As nonlinear functions we tried squashing functions and functions that expanded the input. From several combinations we found that using squashing functions for both nonlinearities yielded the best results.

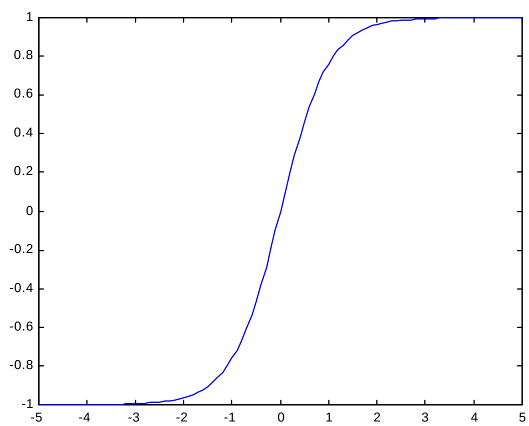

Figure 3: Squashing function used as nonlinear function.

The exact form of the squashing function was not important. In the results presented in next section we used the *tanh*(k(.)) (see figure 3), with a gain k, selected to saturate at half a standard deviation of the input frame.

6. EXPERIMENTS

We compared the LPC gain obtained with one stage (LPC-1), a two stages modeled with the structure proposed in figures 1 and 2 (NLPC-2), and for reference purposes, also a second stage that consisted of an lpc of the same order than the nonlinear stage (LPC-2).

As input speech signal we use the spanish sentence "el golpe de timón fue sobrecogedor", uttered by eight different speakers (four males and four females). The signal, was sampled at 8KHz.

The performance criterion used to evaluate the results is the prediction gain, defined as:

$$Gp(dB) = 10 \log\left(\frac{E\left[x^2(n)\right]}{E\left[e^2(n)\right]}\right)$$

In the first experiment (figure 4) we compare the prediction gain for the first stage (order 12), with the prediction gain when another linear prediction stage of order 12 is used as a second stage, and the gain with the structure proposed in section 4. In order to compare the results, we plot the mean of the prediction gain for the eight speakers, and the margin between one standard deviation for each case. As can be seen, the use of a nonlinear structure yields a significative improvement with respect to the structure of two linear predictors (which are equivalent to a simple linear prediction stage of order 24). In this experiment we used a frame of 256 samples.

The system also was not very sensitive to changes in the length of the filter in the second stage. This is shown in figure 5, where we can see that the prediction gain, in the nonlinear case, degrades slowly as the length is diminished. Also the performance of a nonlinear second stage is better than a linear stage for all the orders considered. The performance degrades for orders lower than 5.

We also studied the effect of the frame length on the prediction gain. The results are summarized in figure 6. It can be seen that for a margin that goes from 256 to 64 samples, the results are consistent, and the structure proposed in this paper outperforms the linear one. It can be seen that the prediction gain is almost the same for all the frame lengths, and in all cases, the confidence margin of the results show that the improvement is significative. Also we can see a slight increasing trend in the nonlinear prediction gain as the frame length diminishes.

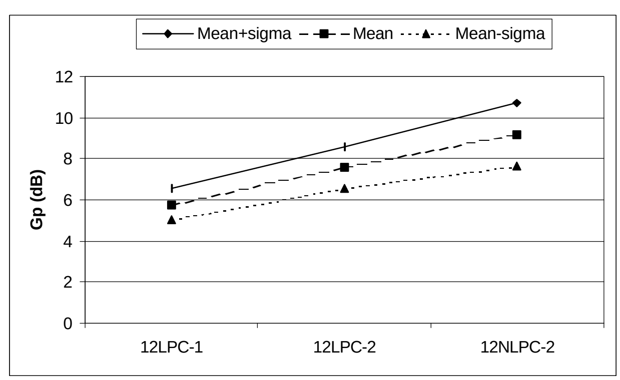

Figure 4: Prediction gain for the first stage (12LPC-1), the first stage and the second linear stage (12LPC-2), and

the first stage and the second nonlinear stage (12NLPC-2).

7. SUMMARY

We have presented a new structure for performing speech linear prediction, which gives better results than the classical LPC methods. The structure is based in two stages, the first removes the linear components of the signal, and the second concentrates in predicting the low energy component of the residual.

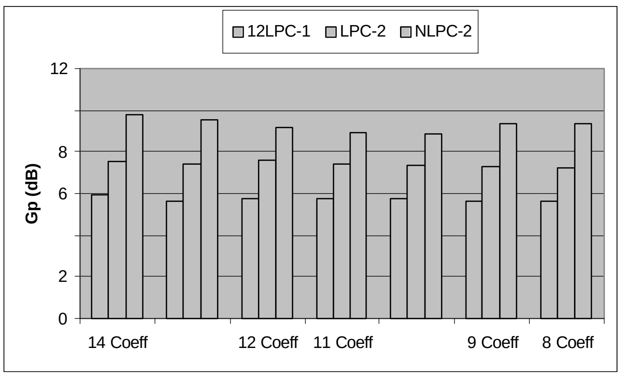

Figure 5: Effect of diminishing the order of the (LP) second filter.

The method is based on a criterion which requires the knowledge of error pdf, or more precisely of the score function. Implicitly, this criterion involves higher order statistics, which can be chosen optimally with a good estimation of the score function, e.g. computed from kernel estimators of the error pdf. Also we present a method for computing the coefficients of a filter between two nonlinearities, for the general case, of arbitrary pdf, and form of the nonlinerities (the only restriction is that they should be invertible).

Real speech signal experiments show that this method is always better than LPC method, for different orders of the second stage, and for different frame lengths.

AKNOWLEDGMENTS

This work has been in part supported by the Direcció General de Recerca de la Generalitat de Catalunya and by the CICYT Spanish research project TIC2000-1005-C03-01.

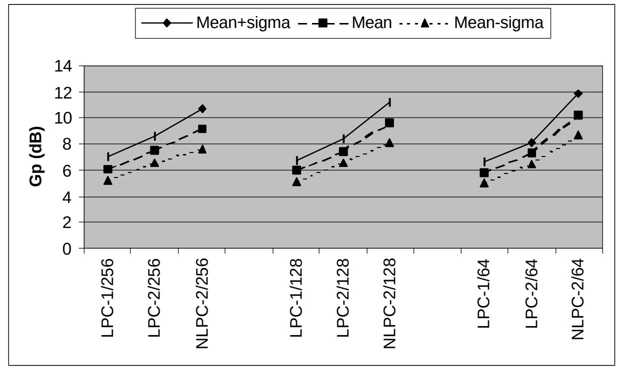

Figure 6: Comparison of the prediction gain for different frame lengths.

8. REFERENCES

[1] A.Taleb, J.Solé, C.Jutten, "Quasi-Nonparametric Blind Inversion of Wiener Systems". *IEEE Transactions on Signal Processing*, vol. 49, nº5, pp. 917-924, May 2001.

[2] G.Kubin, *Nonlinear processing of speech.* Chapter 16, in *Speech coding and synthesis*, W.B. Kleijn & K.K. Paliwal editors, Ed. Elsevier, 1995

[3] M.Faúndez, O.Oliva, "ADPCM with nonlinear prediction", *EUSIPCO'98*, Rhodas, pp. 1205-1208, 1995.

[4] A.Kumar, A. Gersho, "LD-CELP speech coding with nonlinear prediction". *IEEE Signal Processing Letters*, vol. 4, nº4, pp. 89-91, April 1997.

[5] N.Ma, G.Wei, "Speech coding with nonlinear local prediction model", *ICASSP'98*, vol. 2, pp. 1101-1104, 1998.

[6] J.Thyssen, H.Nielsen, S.D.Hansen, "Nonlinear short term prediction in speech coding", *ICASSP'94*, vol. 1, pp. 185-188, April 1994.

[7] J. Solé i Casals , E. Monte i Moreno , C. Jutten , A. Taleb, "Source Separation Techniques Applied To Speech Linear Prediction". *ICSLP2000*, Beijing (China), October 2000.