Stability of Linear Predictive Structures using IIR Filters

A.C. den Brinker
Philips Research, Prof. Holstlaan 4 (WY82), NL-5656 AA Eindhoven, The Netherlands.
Phone: +31 40 27 42802 Fax: +31 40 27 44675
E-mail: bert.den.brinker@philips.com

Abstract—**Considered are general, pure linear prediction schemes, where the prediction of the current sample is based on IIR-filtered past samples. Properties of these schemes are discussed, in particular, the whitening property, the realisation of the synthesis filter and its stability. The realisation problem as occurring in warped linear prediction is solved. Furthermore, it is proved that, at least for a specific class of systems, data input windowing for the calculation of the optimal prediction coefficients guarantees the stability of the synthesis filters.**

Keywords—**LPC, IIR-filters, stability.**

I. Introduction

The use of linear prediction in coding (LPC) has a long history. For the theory, we refer to [1,2]. The conventional linear prediction scheme is based on the tapped-delay-line (TDL) and a prediction of the current sample is made from a linear combination of past samples.

A generalisation of this scheme, called warped linear prediction (WLP), has been proposed. In that case, the tapped-delay-line is replaced by a tapped all-pass line. The combined outputs of a tapped all-pass line can be considered as a realisation of an associated frequency-warped signal. The use of linear prediction in the frequency-warped domain was first introduced in speech coding [3] and was later also applied to audio [4]. By warping, one can make the frequency axis approximately in accordance with a Bark or ERB rate scale [5]. The effect of incorporating warping in the linear prediction is that specific parts of the frequency range are more accurately modelled at the cost of modelling capability in other ranges. For a comparison of WLP and LP we refer to [6].

The warped linear prediction is an example of a scheme which exploits IIR filters to produce the regression signals. However, in contrast to the tapped-delay line, such scheme does not adhere to pure linear prediction: as a result of the all-pass sections, information of the current sample is also present in the regression signals (the outputs of the tapped all-pass line). Consequently, these schemes do not have the whitening property associated with the TDL prediction scheme. Furthermore, either the synthesis filter has to be different from the analysis filter in order to avoid delay-free loops [3] or a two-step processing procedure has to be used [7].

Here we consider general prediction schemes based on one-time delayed, causal, stable IIR-filters. The restriction to IIR filters having no direct feed-through ensures that the predictor makes an estimate of the current sample based on knowledge of past samples of the signal only. We will refer to this case as pure linear prediction.

A natural question is how far properties associated with linear prediction using a transversal filter (tapped-delay-line) generalise to this situation. In particular, the following three conclusions are drawn. The absence of direct feed-throughs gives the possibility of direct realisation of the synthesis filter. It is found that, at least for a specific class of IIR structures, the whitening property holds and the stability of the synthesis filter is guaranteed if the prediction coefficients are calculated using data input windowing. Examples of interesting structures for audio and speech processing are discussed.

II. IIR-based linear prediction schemes

Considered is the following linear prediction scheme. We have a signal called x and a set of K signals y_k, $k = 1, \cdots, K$. We make a linear prediction $\hat{x}$ of x from the regressor signals y_k by

$$\hat{x} = \sum_{k=1}^{K} \alpha_k y_k \tag{1}$$

where α_k are the prediction coefficients. These parameters are defined and calculated according to some optimisation criterion, typically least squares.

The regression signals can be taken as filtered versions of the signal x, thus

$$Y_k = G_k X \tag{2}$$

where X and Y_k are the z-transforms of x and y_k and G_k is the kth transfer function. The conventional situation (called forward linear prediction or one-step-ahead predic-

© STW, 2001

tion) occurs if

$$y_k(n) = x(n-k) \tag{3}$$

or, writing it as filtering operations,

$$G_k(z) = z^{-k}. \tag{4}$$

For this situation we know that if the minimisation criterion is defined by minimum mean squares, the system whitens the residual signal. Furthermore, calculation of the prediction coefficients using data input windowing (autocorrelation method [1,2]) ensures stability of the synthesis filter.

As a generalisation of this scheme we have

$$G_k(z) = \left(\frac{-\lambda + z^{-1}}{1-\lambda z^{-1}}\right)^k \tag{5}$$

with $-1 < \lambda < 1$, i.e., the filtering scheme can be implemented as a tapped all-pass line. We note that for $\lambda = 0$, the tapped-delay-line reappears. As mentioned in the introduction, this type of processing is associated with standard linear prediction in the frequency-warped domain and, hence, is called warped linear prediction [3, 8]. Frequency warping is an interesting possibility for controlling the time-frequency resolution.

Warped linear prediction has been studied extensively [3, 4, 8–11] and we mention the following properties. Using the minimum mean squared prediction error to determine the prediction coefficients does not result in whitening of the residual signal. Furthermore, for data input windowing, an autocorrelation kind of method can be defined leading to stable synthesis filters and normal equations involving a Toeplitz matrix. However, for the synthesis filter we have to deal with a problem. Either, this filter is not directly realisable because of the delay-free loops [3], or the processing in the synthesis filter has to be done by a two-step procedure [7].

In this paper we discuss the situation where the filtering operations G_k are restricted to

$$G_k(z) = z^{-1} H_k(z) \tag{6}$$

with H_k stable and causal filters, in particular IIR-filters. Since now the filters G_k have no direct feed-through, the prediction of the current sample is based on knowledge of past samples only. We will refer to this as pure linear prediction. The analysis scheme is shown in Fig. 1. As a consequence of the choice (6), the synthesis filter is directly realisable in a form similar to the analysis filter since no delay-free loops occur; see Fig. 1.

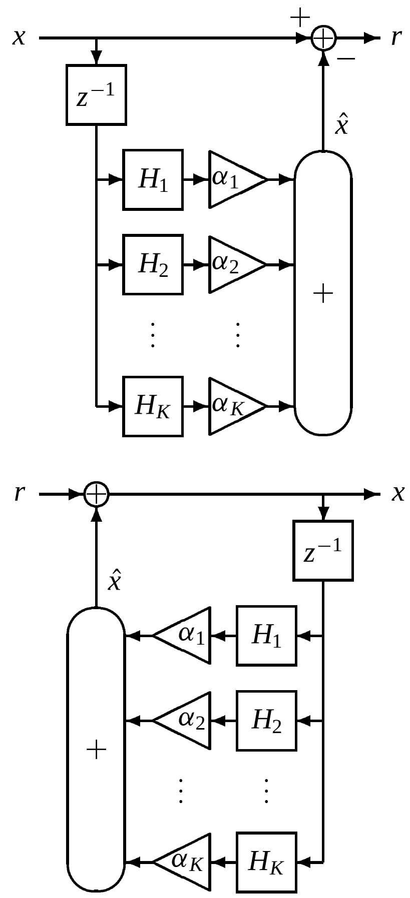

Fig. 1. Linear prediction scheme. Top: analysis filter, bottom: synthesis filter. The filters H_k are stable and causal. The input signal is x, the prediction of x is called $\hat{x}$. The prediction coefficients are α_k and residual signal is r.

III. Prediction coefficients

As criterion J we take the deterministic measure

$$J = \sum_{n=-\infty}^{\infty} |r(n)|^2 \tag{7}$$

with $r = x - \hat{x}$. In practice, x is derived from a signal s by windowing: $x(n) = s(n)w(n)$. Typically, a series of (overlapping) windows is used to find the evolution of the prediction coefficients over time. We also note that, as a consequence of taking the error over the entire time axis, all properties have a counterpart in the stochastic theory; this follows directly from replacing time averages by ensemble averages.

The minimum value of J that can be achieved is called $\hat{J}$ and the values of α for which this is attained are denoted by $\hat{\alpha}$:

$$\hat{J} = \min_{\alpha} J, \tag{8}$$

$$\hat{\alpha} = \arg\min_{\alpha} J. \tag{9}$$

Since r is a linear function of the prediction coefficients, we are dealing with a convex optimisation problem having a single optimum or an underdetermined set of equations. The solution of the problem is given by the normal equations which read in matrix form

$$Q\hat{\alpha} = P \tag{10}$$

where the matrix Q and the vector P are given by their elements

$$Q_{k,l} = \sum_n y_l(n) y_k^*(n), \tag{11}$$

$$P_k = \sum_n x(n) y_k^*(n), \tag{12}$$

where * denotes complex-conjugation and $1 \leq k \leq K$. We note that the matrix Q is a Hermitian, positive semi-definite matrix. In order that there is a unique solution, the signals y_k have to be linearly independent, which immediately translates into the necessary condition that the set of transfer functions G_k is linearly independent. Of course, linear independence of the signals y_k also translates into a condition on the windowed input signal x. For instance, the signal $x = 0$ is forbidden. In practice, there are several ways to deal with such situations.

IV. Stability and whitening

For practical significance of the proposed linear prediction scheme, the stability of the synthesis filter is obviously an issue. What we will show is that the input-windowed optimisation criterion yields stable synthesis filters at least for a restricted class of filters H_k.

The total transfer $x \to r$ can be written as

$$F(z) = 1 - z^{-1} \sum_{k=1}^{K} \alpha_k H_k(z) = \frac{N(z)}{D(z)} \tag{13}$$

where N and D are the numerator and denominator polynomials. For N we write

$$N(z) = 1 - \sum_{n=1}^{L} b_n z^{-n}. \tag{14}$$

In general, L is larger than $K-1$ and less or equal to the sum of the orders O_k of H_k plus 1:

$$K \leq L \leq 1 + \sum_{k=1}^{K} O_k.$$

Since there are K degrees of freedom in the coefficients α_k, the polynomial coefficients in N are subject to restrictions. More specifically, the polynomial coefficients b_n adhere to a set of $L-K$ linear constraints. For the specific situation that $K = L$ we have the following theorem.

Theorem.
If a unique solution exists and if the number of degrees of freedom K of the analysis filter equals the order L of the numerator polynomial of F, then the optimal filter F according to the minimisation of $J = \sum_n |r(n)|^2$ is a minimum-phase filter.

Proof.
Consider the filter F_1 with

$$F_1(z) = \frac{\prod_{k=1}^{K}(1-\beta_k z^{-1})}{D(z)}$$

and where $\beta_K \geq \beta_{K-1} \geq \cdots \beta_1$. Suppose that F_1 is a non-minimum phase filter, thus $|\beta_K| > 1$. We define

$$F_2(z) = \frac{(1 - z^{-1}/\beta_K^*) \prod_{k=1}^{K-1}(1-\beta_k z^{-1})}{D(z)},$$

i.e., we change the pole β_K to a pole inside the unit circle. For the error criterion J we have

$$\begin{aligned} J &= \frac{1}{2\pi} \int_{<2\pi>} |X(e^{j\theta})|^2 |F_1(e^{j\theta})|^2 d\theta \\ &= |\beta_K|^2 \frac{1}{2\pi} \int_{<2\pi>} |X(e^{j\theta})|^2 |F_2(e^{j\theta})|^2 d\theta \\ &> \frac{1}{2\pi} \int_{<2\pi>} |X(e^{j\theta})|^2 |F_2(e^{j\theta})|^2 d\theta, \end{aligned}$$

where X is the z-transform of the input signal x and assuming that $X \neq 0$. Thus, a filter F_1 where the zero with largest radius lies outside the unit circle can never attain the minimum of J.

We note the following.
1. The restriction $L = K$ is a severe one. But even with this restriction, a set of interesting IIR-based LPC schemes evolves; see V.
2. There are many more situations where this algorithm yields stable synthesis filter; the restrictions on H_k are certainly much milder than given here.
3. If the analysis filter F is minimum phase, then the synthesis filter is (marginally) stable.
4. The proof given here also works when considering the stochastic setting.
5. The given theorem holds for the TDL as well, the proof is simple and differs from that in textbooks [1,2].

Since the optimal filter is a minimum phase filter, it follows that

$$\frac{1}{2\pi} \int_{<2\pi>} \log |F(e^{j\theta})|^2 d\theta = 0, \tag{15}$$

just like in the TDL case. Therefore, the degree of freedom resides in the shape of the frequency characteristic and the minimum of J is attained by maximum flatness of the spectrum of r.

V. Examples

Some examples of IIR structures adhering to the condition for which stability is guaranteed are given below.

Laguerre filters [12, 13]

$$H_k(z) = \frac{\sqrt{1-|\lambda|^2}}{1-z^{-1}\lambda}\left\{\frac{z^{-1}-\lambda}{1-z^{-1}\lambda}\right\}^{k-1}$$

with $\lambda \in (-1, 1)$. Note that TDL is a special case of Laguerre. The effect of λ is exactly that as in WLP: a shift of modelling capability to either the low ($\lambda > 0$) or high frequencies ($\lambda < 0$).

Kautz filters [13, 14]

$$H_k(z) = \frac{\sqrt{1-|\lambda_k|^2}}{1-z^{-1}\lambda_k}\prod_{m=1}^{k-1}\frac{z^{-1}-\lambda_m^*}{1-z^{-1}\lambda_m}$$

with $\lambda_m \in \mathbb{C}$ and $|\lambda_m| < 1$. Note that Laguerre is a special case of Kautz. We note that the Kautz filters have much more freedom in a priori tuning. For instance, different complex λ_ks correspond to bandpass filters. In this way, the frequency axis can be subdivided into arbitrary unequal regions.

Combinations of first-order sections

$$H_k(z) = \sum_{l=1}^{K}\beta_{k,l}B_l(z) ,$$

where $\beta_{k,l}$ are entries from an invertible matrix β and the partial transfers $B_l(z)$ are the Kautz filters

$$B_l(z) = \frac{\sqrt{1-|\lambda_l|^2}}{1-z^{-1}\lambda_l}\prod_{m=1}^{l-1}\frac{z^{-1}-\lambda_m^*}{1-z^{-1}\lambda_m}$$

with $\lambda_m \in \mathbb{C}$ and $|\lambda_m| < 1$. The λ_m's are not necessarily distinct.

Since the Laguerre filters are basically identical to the allpass-line as is used in WLP, it is expected that this structure exhibits a behaviour similar to that of WLP. This assertion has been tested and the results can be found in more detail in a companion paper in this proceedings [15]. Furthermore, it is shown there that Kautz filters with proper parameterisation behave similarly as well. As already stated in the Introduction, both structures overcome the realisation problem of the synthesis filter as occurs in WLP.

VI. Discussion

We have considered a linear prediction scheme, called pure linear prediction, where the prediction of the current sample is based on IIR-filtered past samples. Such a scheme generalises the conventional tapped-delay-line but differs from the generalisation known as warped linear prediction. The realisation problem as occurring in warped linear prediction is solved. Furthermore, it is proved that, at least for a specific class of systems, data input windowing for the calculation of the optimal prediction coefficients guarantees the stability of the synthesis filters. Kautz and Laguerre filters are structures belonging to this class.

References

[1] J.D. Markel and A.H. Gray Jr. *Linear prediction of speech*. Springer Verlag, Berlin, 1976.

[2] S. Haykin. *Adaptive filter theory*. Prentice Hall, Englewood Cliffs, N.J., USA, 1991.

[3] H.W. Strube. Linear prediction on a warped frequency scale. *J. Acoust. Soc. Am.*, 68:1071–1076, 1980.

[4] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U.K. Laine, and J. Huopaniemi. Frequency-warped signal processing for audio applications. *J. Audio Eng. Soc.*, 48:1011–1031, 2000.

[5] J.O. Smith and J.S. Abel. Bark and ERB bilinear transform. *IEEE Trans. Speech Audio Process.*, 7:697–708, 1999.

[6] A. Härmä and U.K. Laine. A comparison of warped and conventional linear prediction. *IEEE Trans. Speech and Audio Process.*, 9:589–597, 2001.

[7] A. Härmä. Implementation of frequency-warped recursive filters. *Signal Process.*, 80:543–548, 2000.

[8] U.K. Laine, M. Karjalainen, and T. Altosaar. WLP in speech and audio processing. In *Proc. ICASSP-94*, volume II, pages 349–352, Adelaide, Australia, 1994.

[9] A. Härmä, U.K. Laine, and M. Karjalainen. Warped linear prediction (WLP) in audio coding. In *Proc. NorSig-96*, pages 447–450, Espoo, Finland, 24-27 Sept. 1996.

[10] A. Härmä, U.K. Laine, and M. Karjalainen. WLPAC – A perceptual audio codec in a nutshell. AES 102nd Conv., Munich, Germany, 1997. Preprint 4420.

[11] B. Edler and G. Schuller. Audio coding using a psycho-acoustic pre- and post-filter. In *Proc. ICASSP2000*, pages 1881–1884, Istanbul, Turkey, 5–9 June 2000.

[12] Y.W. Lee. Synthesis of electrical networks by means of Fourier transforms of Laguerre functions. *J. Math. Phys.*, 11:83–113, 1932.

[13] P.W. Broome. Discrete orthonormal sequences. *J. Association for Computing Machinery*, 12:151–165, 1965.

[14] W.H. Kautz. Transient synthesis in the time domain. *IEEE Trans. Circuit Theory*, 1:29–39, 1954.

[15] V. Voitishchuk, A.C. den Brinker, and S.J.L. van Eijndhoven. Alternatives for warped linear predictors. In *Proc. ProRISC Workshop CSSP*, Veldhoven (NL), 29-30 Nov. 2001. This issue.