Similarities and differences between Warped Linear Prediction and Laguerre Linear Prediction

Citation for published version (APA):
Brinker, den, A. C., Krishnamoorthi, H., & Verbitskiy, E. A. (2011). Similarities and differences between Warped Linear Prediction and Laguerre Linear Prediction. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(1), 24-33. https://doi.org/10.1109/TASL.2010.2042130

DOI:
10.1109/TASL.2010.2042130

Document status and date:
Published: 01/01/2011

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Similarities and Differences Between Warped Linear Prediction and Laguerre Linear Prediction

Albertus C. den Brinker, *Senior Member, IEEE*, Harish Krishnamoorthi, *Student Member, IEEE*, and Evgeny A. Verbitskiy

***Abstract*—Linear prediction has been successfully applied in many speech and audio processing systems. This paper presents the similarities and differences between two classes of linear prediction schemes, namely, Warped Linear Prediction (WLP) and Laguerre Linear Prediction (LLP). It is shown that both systems are closely related. In particular, we show that the LLP is in fact a WLP system where the optimization procedure is adapted such that the whitening property is automatically incorporated. The adaptation consists of a new linear constraint on the parameters. Furthermore, we show that an optimized WLP scheme where whitening is achieved by prefiltering before estimating the optimal coefficients results in a filter having all except the last reflection coefficient equal to those of the optimal LLP filter.**

***Index Terms*—Audio coding, frequency warping, linear prediction, speech coding.**

I. Introduction

LINEAR prediction is a simple and popular technique used in the coding of speech signals. Here, an input signal is modeled such that the current sample is predicted from a linear combination of past samples [1]. Usually, a mean-squared-error optimization criterion is used to define the optimal predictor parameters, which results in the well-known Yule–Walker equations. Moreover, the technique of linear prediction is associated with a number of desirable properties that can be of benefit in many applications. For example, the reflection coefficients that are obtained as a by-product of solving the normal equations ensure simple control of the stability of the synthesis filter when quantizing these parameters. Additionally, the whitening property associated with the minimization process ensures a spectrally flat error signal. This implies that the error signal is restricted to a particular class of signals, and this knowledge can be exploited in coding by constructing an appropriate code book. A comprehensive overview of linear prediction can be found in [2] and [3].

Several variants of linear prediction based on warped signal processing concepts [4], such as WLP [5]–[7] and LLP [8]–[11] have been reported. The primary motivation behind employing warped processing is its ability to process acoustic signals according to the frequency resolution of the human auditory system [12].

This paper aims at clarifying the relations between these different systems. Section II introduces warping, the two known variants of WLP (WLP-A and WLP-B), and the Laguerre linear prediction system (LLP). Section III gives some experimental observations, leading to the conclusion that the WLP-A system has to be very closely related to the LLP system. This is further explored from a theoretical point of view. Section IV shows that the LLP system is in fact a third variant of WLP where the optimization corresponds to minimization of the output signal energy of a warped predictor under a linear constraint on the parameters that ensures whitening. Section V proofs that for systems of order p, the first $p-1$ reflection coefficients of the WLP-A and LLP systems are identical. The last section contains the conclusions.

As a vehicle to compare the two adaptive filter systems, we use minimization of the output power as the optimization criterion. There are many other ways of defining an optimal filter, e.g., discrete all-pole modeling [13], minimum variance distortionless response [14], or least absolute error. A comparison of some of these criteria can be found in [15]. In this paper, we stick to minimum output power since this is mathematically tractable and we will argue that the conclusions that we draw from this specific choice carry over to other optimization criteria.

II. Linear Prediction Based on Warping

A. Frequency Warping

As a formal definition of a warping function which is broad enough for the current purpose we will use the following.

Definition: A function $\phi(\theta)$ is a warping function if it is a continuous, monotonically increasing function mapping the interval $(-\pi,\pi]$ onto itself.

A very convenient warping function is given by

$$e^{-\jmath\phi(\theta)} = \frac{-\lambda + e^{-\jmath\theta}}{1-\lambda e^{-\jmath\theta}} \tag{1}$$

with $-1 < \lambda < 1$ and $\jmath = \sqrt{-1}$. The convenience stems from the fact that it is related to a realizable filter: a first-order allpass section. We denote this specific warping as $\mathcal{W}$.

A frequency-warped signal can now be defined as follows.

Manuscript received April 24, 2009; revised August 24, 2009; accepted January 13, 2010. Date of publication March 15, 2010; date of current version October 01, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrick A. Naylor.

A. C. den Brinker is with Philips Research, NL-5656 AE Eindhoven, The Netherlands (e-mail: bert.den.brinker@philips.com).

H. Krishnamoorthi was with the Signal Processing Group, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands. He is now with Arizona State University, Tempe, AZ 85287 USA (e-mail: harish.krishnamoorthi@asu.edu).

E. A. Verbitskiy was with Philips Research, NL-5656 AE Eindhoven, The Netherlands. He is currently with the Mathematical Institute, Leiden University, 2300 RA Leiden, The Netherlands (e-mail: evgeny@math.leidenuniv.nl).

Digital Object Identifier 10.1109/TASL.2010.2042130

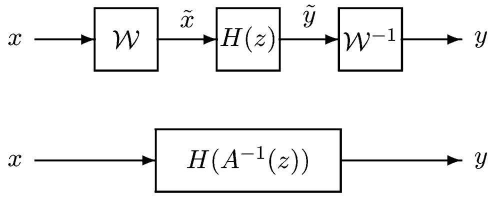

Fig. 1. Equivalence between frequency warping $\mathcal{W}$, time-invariant linear processing H, and de-warping $\mathcal{W}^{-1}$ and warped time-invariant linear processing.

Definition: Suppose x is a signal with z-transform X. The signal $\tilde{x}$ with z-transform $\tilde{X}$ is the frequency-warped signal with warping function $\phi(\theta)$ if $X(e^{-\jmath\theta}) = \tilde{X}(e^{-\jmath\phi(\theta)})$.

Determining a warped signal $\tilde{x}$ from x is not very practical: in principle, one needs to know the entire signal from $-\infty$ to $+\infty$. However, one can make a warped signal from a causal signal as described in [4]. There, the warped signal is obtained by propagating an input signal through a chain of first-order allpass filters preceded by certain prefilters. The pole-zero location associated with the allpass filter can be set to obtain the desired frequency mapping.

A much easier thing is to apply warping to processing. Consider the following setup. First, we warp the signal: $x \to \tilde{x}$. Next, we filter the warped signal by a linear time-invariant system with transfer function $H(z)$ which produces the output signal $\tilde{y}$. Lastly, we perform an inverse warping on $\tilde{y}$ to obtain y. We have

$$Y(e^{\jmath\theta}) = \tilde{Y}(e^{\jmath\phi}) = \tilde{X}(e^{\jmath\phi})H(e^{\jmath\phi}) = H\left(e^{\jmath\phi(\theta)}\right)X(e^{\jmath\theta}). \tag{2}$$

In particular, if we are warping according to the warping function (1), it means that we can directly absorb the warping and de-warping into the filter operation by replacing in the filter $H(z)$ all delay operators z^{-1} by a first-order allpass section $A(z)$ with

$$A(z) = \frac{-\lambda + z^{-1}}{1 - \lambda z^{-1}}. \tag{3}$$

This idea is also shown in Fig. 1.

The possibility of incorporating the warping and de-warping into the processor block in the middle (as shown in Fig. 1) holds for linear time-invariant systems. If, however, the middle block is a nonlinear or time-variant system, this approach of replacing the delays z^{-1} from the middle block by allpasses $A(z)$ does not, in general, lead to identical behavior. This also holds for adaptive filtering. In that case, we are working with stochastic signals described by power spectral density functions. The warping changes the frequency axis and therefore changes the shape of the density function as well. We will see this effect in warped linear prediction (next section) in the form that the equivalence between error energy minimization and whitening (as we know it from conventional linear prediction) no longer holds.

B. Warped Linear Prediction

The idea of linear prediction on a warped frequency scale was first introduced by Strube in [5]. Here, the unit-delay elements z^{-1} in the conventional predictor structure are replaced with allpass sections $A(z)$ with

$$A(z) = \frac{-\lambda + z^{-1}}{1 - \lambda z^{-1}} \tag{4}$$

where the parameter λ ($|\lambda| < 1$) can be chosen to obtain the desired frequency warping. Hence, the warped linear prediction, $\hat{x}(n)$, of $x(n)$ is given by

$$\hat{x}(n) = -\sum_{k=1}^{p} a_k x_k(n) \tag{5}$$

where $x_k(n) = h_k(n) * x(n)$, $h_k(n)$ represents the inverse z-transform of $A^k(z)$, the a_k's represent the filter coefficients, and "$*$" denotes the convolution operation. The error signal $e(n)$ is obtained as

$$e(n) = x(n) - \hat{x}(n). \tag{6}$$

The prediction error filtering $x \to e$ can be expressed as

$$F(z) = 1 + \sum_{k=1}^{p} a_k A^k(z). \tag{7}$$

The optimal parameters $\hat{a}_k$ for the predictor of (5) can be found in a number of different ways. Typically, a mean-squared error criterion is taken to determine the $\hat{a}_k$'s. The mean-squared error ϵ can be formally expressed as

$$\epsilon = E\left[\left(x(n) + \sum_{k=1}^{p} a_k x_k(n)\right)^2\right] \tag{8}$$

where E denotes the expectation operator and ϵ represents the residual error energy. We assume a wide-sense stationary signal x in which case ϵ does not depend on time.

Minimization of (8) leads to the following set of equations:

$$\sum_{k=1}^{p} \hat{a}_k E\left[x_k(n)x_i(n)\right] = -E\left[x(n)x_i(n)\right], \text{ for } 1 \leq i \leq p. \tag{9}$$

Equation (9) is widely referred to as the normal or Yule–Walker (YW) equations and represents a set of p linear equations in p unknowns. It relates an autocorrelation sequence $\rho(k) = E[x(n)x_k(n)]$, $k = 0, \cdots, p$ to a minimum-phase filter defined by the coefficient sequence $\hat{a}_k$, $k = 0, \cdots, p$ (with $\hat{a}_0 = 1$).

Eq. (9) can also be symbolically represented as

$$R\underline{\hat{a}} = -\underline{q} \tag{10}$$

where R is a $p \times p$ autocorrelation matrix, $\underline{\hat{a}} = [\hat{a}_1\ \hat{a}_2\ \ldots\ \hat{a}_p]^T$, and $\underline{q}$ are $p \times 1$ vectors. It can be easily verified that the matrix R is symmetric and Toeplitz (see the Appendix) and that the right-hand side vector $\underline{q}$ is structurally related to the matrix R. For solving this type of equation, efficient algorithms such as the Levinson–Durbin algorithm [16] can be employed. In addition to the filter coefficients, reflection coefficients are obtained during the recursive solution in the Levinson–Durbin algorithm. The minimum-phase property of the filter restricts the reflection coefficients to have an absolute value less than 1. The reflection coefficients can also be found from the polynomial coefficients $\hat{a}_k$ by the backward recursion algorithm [2]. We also note that the kth reflection coefficient is uniquely defined by $\rho(i)$ with $i = 0, \cdots, k$.

Although the steps involved in warped linear prediction scheme are similar to the conventional linear prediction scheme, several differences exist between the two schemes. In the next subsection, differences in terms of the spectral characterization and synthesis filter realizations are presented.

1) Spectral Characterization: For the warped linear predictor, a mean-squared error minimization procedure results in the set of normal equations as described in (9). Here, the autocorrelation terms are obtained from $\rho(i-j) = E[x_i(n)x_j(n)]$. The index in ρ refers to the difference in number of allpass sections used between x_i and x_j and does not, as in conventional linear prediction, refer to a time-lag. A consequence is that the mean-squared error criterion (8) minimizes the error on the warped frequency axis. Therefore, the resultant residual error energy is also whitened on the warped frequency axis [5]. In many cases, it is preferred to have all correlations removed in the residual, i.e., to have a flat spectrum for the output signal. Two techniques have been proposed in the literature to achieve this.

WLP-A In [5], a prefilter, $W(z) = \sqrt{1-\lambda^2}/(1-\lambda z^{-1})$, is introduced during the minimization procedure. This results in a new set of normal equations that are now solved to obtain the optimal coefficients. The optimal coefficients defined in this way are denoted as α_k, $k = 1, \cdots, p$. The predictor filter formed with these coefficients now attains minimum error energy and spectral flatness. We stress that the $W(z)$ prefilter is only employed during the minimization process and is not present in the actual predictor filter. This means we have signals $y_k(n)$ defined as $y_k = x * g_k$ with the z-transform G_k of g_k being

$$G_k(z) = W(z)A^k(z) \text{ for } k = 0, \cdots, p \qquad (11)$$

being used in the optimization. The optimal coefficients are defined according to (9) but with the replacement of signals x and x_k by y_0 and y_k, respectively. The WLP filter that is actually used, however, still uses the signals x and x_k to produce the output signal. We will call this output signal the residual (i.e., not the error signal which is minimized) and denote it as $\tilde{e}_A$. This is depicted in Fig. 2.

WLP-B As an alternative to prefiltering, the input signal before determining the optimal coefficients, it is also possible to apply a postfilter $W^{-1}(z) = (1-\lambda z^{-1})/\sqrt{1-\lambda^2}$ on the error signal in order to obtain a spectrally flat signal [5], [6]. Thus, we use the optimal coefficients $\hat{a}_k$ as defined by the YW equations (9) and add a postfilter; this is depicted in Fig. 3.

We note that both systems do not directly minimize the output error signal on the normal frequency scale. Instead, this is mimicked by the pre- or postfilter. In that sense they are presumably suboptimal to a system which is inherently whitening and has the output power as optimization target; this is shown in

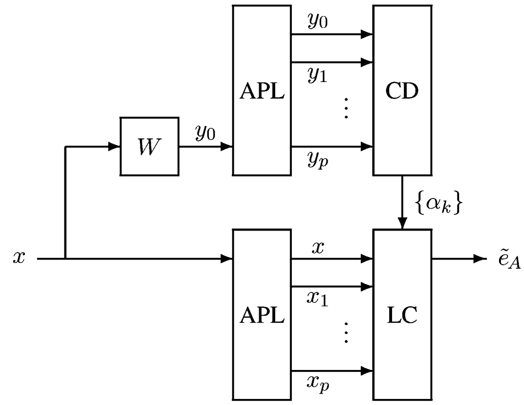

Fig. 2. WLP-A scheme consisting of the prefilter W, two allpass lines (APL), a coefficient determiner (CD), and a linear combiner (LC).

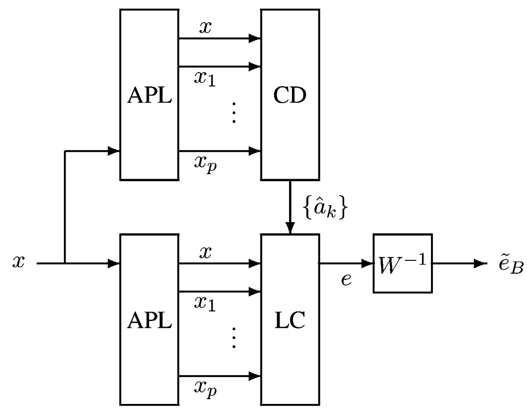

Fig. 3. WLP-B scheme consisting of two allpass lines (APL), a coefficient determiner (CD), a linear combiner (LC), and the postfilter W^{-1}.

Section III. We also note that both approaches do not yield exactly the same filter. In the first case, the designed filter is of the form

$$F_A(z) = 1 + \sum_{k=1}^{p} \alpha_k A^k(z) \qquad (12)$$

while in the second case we have

$$F_B(z) = \frac{1-\lambda z^{-1}}{\sqrt{1-\lambda^2}} \left(1 + \sum_{k=1}^{p} \hat{a}_k A^k(z)\right). \qquad (13)$$

In general, given a set $\{\hat{a}_i\}$, there does not exist a set $\{\alpha_i\}$ such that $F_A(z)$ and $F_B(z)$ are equal.

Note also that in case that the input signal is a white signal, we have $\alpha_k = 0$ (i.e., for WLP-A). This is obviously not the case for the second procedure (WLP-B); there we have that the normalized sequence $\rho(k)$ equals $(-\lambda)^k$ and thus that the optimal predictor coefficients are $\hat{a}_1 = \lambda$, $\hat{a}_k = 0$ for $k = 2, \cdots, p$. In fact, now the $\hat{a}_k$ are actually used to compensate the postfilter.

2) Synthesis Filter Realizations: The behavior of the synthesis filter represents another important difference between the conventional linear prediction and warped linear prediction schemes. The transfer function of the synthesis filter is obtained by taking the reciprocal of the transfer function of the analysis filter. In the warped linear prediction scheme, the allpass sections in the analysis filter introduce delay-free loops in the

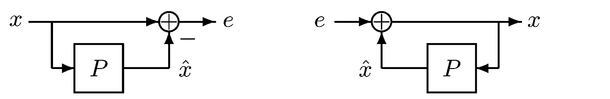

Fig. 4. Analysis and synthesis filter realization with the predictor P in a feedforward and feedback loop, respectively.

synthesis filter. Therefore, the synthesis filters are not directly realizable. To overcome this limitation, Strube proposed an alternate filter structure for the synthesis filter that avoids the delay-free loops [5]. Furthermore, he developed a mapping procedure to obtain the coefficients associated with the alternate filter structure from that of the predictor filter. Although this mapping procedure overcomes the issue of delay-free loops, it demands additional computational complexity and is ill-conditioned [8]. This issue was further addressed in [17] where two techniques are considered. The first technique consists of switching to a different filter structure such that the delay-free loops are eliminated. The second technique concerns direct implementation of the delay-free loops. Although in the latter case, the predictor structure in encoder and decoder can be identical, this is not true for the signals within the network. In both cases, the straightforward predictor implementation $P(z) = -\sum_{k=1}^{p} a_k A^k(z)$ typically used in the analysis filter and shown in Fig. 4 can not be maintained in the same form in the synthesis filter. Only if the predictor P is a cascade of a delay z^{-1} and a causal second filter, then we can realize the predictors in the encoder and decoder as shown in Fig. 4 while guaranteeing exact equal behavior (i.e., including identical states, identical signals at corresponding nodes and identical multipliers in the analysis and synthesis predictor P). This is important for perfect reconstruction (e.g., lossless coding) in actual implementations where finite word-length arithmetic is used.

C. Pure Linear Prediction

Pure linear prediction considers prediction of an input signal from its infinite impulse response (IIR) filtered versions of one sample delayed input signal [8]. This scheme is associated with a number of desirable properties; it ensures 1) spectral flatness of the residual signal, and 2) the prediction filter in the analysis and synthesis filters can be taken identically. In [8], a class of filter transfer functions for which stability of synthesis filters is guaranteed is further highlighted. The set of discrete Laguerre functions [18], [19] is one such example that belongs to this class. The transfer function of the Laguerre-based prediction (LLP) scheme can be expressed as

$$F_L(z) = 1 + z^{-1} \sum_{k=1}^{p} b_k \frac{\sqrt{1-\lambda^2}}{1-\lambda z^{-1}} \left(\frac{-\lambda + z^{-1}}{1-\lambda z^{-1}} \right)^{k-1}. \quad (14)$$

The Laguerre-based pure linear prediction scheme combines the advantages associated with both warped and conventional linear prediction schemes. The optimal coefficients for b_k are denoted as $\beta_k, k = 1, \cdots, p$ and are defined by minimum mean-squared error of the output signal. More details are provided later on in this paper.

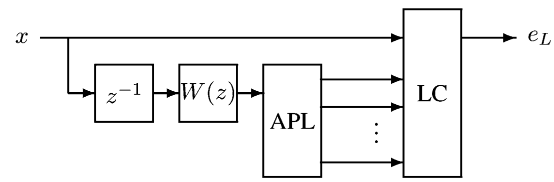

Fig. 5. Analysis filter in the LLP system.

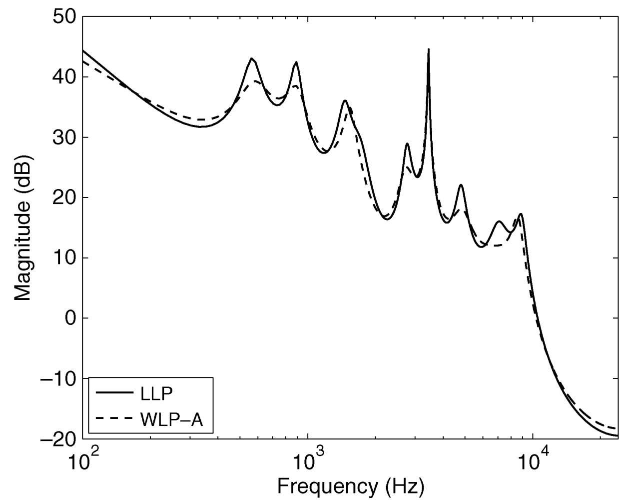

Fig. 6. Example of the amplitude responses of the synthesis filters of a WLP-A and an LLP system. The order was set to $p = 20$.

III. Problem Statement

In this section, several differences and similarities between the WLP and the Laguerre Linear Prediction (LLP) scheme are given. In particular, we compare the WLP-A with the LLP system. The WLP-A and LLP schemes are illustrated in Figs. 2 and 5, respectively. The results of these comparisons is what actually motivated us to consider the systems more closely from a mathematical point of view as is done in Sections IV and V.

A. Transfer Functions

To start with, the analysis filters are different. We have

$$F_A(z) = 1 + \sum_{k=1}^{p} \alpha_k A^k(z) \quad (15)$$

and

$$F_L(z) = 1 + z^{-1} \sum_{k=1}^{p} \beta_k \frac{\sqrt{1-\lambda^2}}{1-\lambda z^{-1}} \left(A(z)\right)^{k-1}. \quad (16)$$

Therefore, we expect that both systems have a different transfer characteristic even though the optimization is defined in a similar way, namely minimal energy and a spectrally flat output signal of the filter. In practice however, the transfer functions are nearly identical. This is shown in the following example.

In Fig. 6, we have plotted an example of the amplitude characteristics of the transfer functions of the synthesis filters of the WLP-A and LLP system. The order of both systems p was set to 20. The input was a signal sampled at 48 kHz. To calculate the optimal parameters, we used a pole $\lambda = 0.7$ and segments of 1024 samples (≈21 ms), which are windowed by a Hanning window. We observe a close match between the two responses.

In general, the WLP-A system delivers slightly smoother responses (slightly less pronounced peaks). The differences decrease with increasing order.

We note that, in order to facilitate the comparison, we recalculated the transfer function of the WLP-A system to that having a mean 0-dB amplitude transfer. Later on it is shown that this can be achieved by dividing the transfer function $F_A(z)$ by D with

$$D = \sum_{k=0}^{p} \alpha_k (-\lambda)^k \tag{17}$$

with $\alpha_0 = 1$ (see also [6]). Furthermore, we note that $\lambda = 0.7$ which is close to the optimal warping factor for modeling the frequency resolution of the human auditory system [12]. For the purposes of revealing similarities and differences, the exact value of λ is irrelevant as long as it is not 0 since then we return to conventional linear prediction case. Our results will carry over to any other λ as will become clear from the theoretical analysis.

B. Spectral Flatness

The question that the previous example may raise is how general the conclusions drawn from a particular example are. In order to get more grip on the issue of difference in spectral response, we consider the spectral flatness measure of the error signals. We take the definition of the spectral flatness measure γ [3] as

$$\gamma^2 = \frac{\exp\left\{\frac{1}{2\pi}\int \ln P(\theta)d\theta\right\}}{\frac{1}{2\pi}\int P(\theta)d\theta} \tag{18}$$

where P is the power spectral density function of the considered signal and the integral is taken over the interval $[0, 2\pi)$.

We consider the residual signals from the WLP-A system, scaled WLP-A system and LLP system and denote these as e_w, $e'_w = e_w/D$, and e_l. The associated spectral flatness measures are denoted as γ_w, γ'_w, and γ_l. We note that the spectral flatness is independent of amplitude scaling and thus $\gamma_w = \gamma'_w$. Furthermore, it is easy to show (see the Appendix) that

$$\frac{\gamma_l^2}{\gamma_w^2} = \frac{\epsilon_w/D^2}{\epsilon_l} \tag{19}$$

where ϵ_w and ϵ_l are the output signal powers of warped and Laguerre system, respectively. Thus, the ratio of the output powers immediately reflects the ratio of the residual spectral flatness measures.

C. Output Power

Since warping has been proposed as a tool for full-band audio coding, we took a collection of short excerpts containing music and speech to measure the output powers of the WLP-A system and the Laguerre system. In total, we used 43 excerpts sampled at 48 kHz, each excerpt of about 10-s duration. We took segments of 1024 samples ($\approx$21 ms) with an update of 512 samples.

In Fig. 7, we have plotted the power difference (in dB) of the residuals in the form of a histogram for two different prediction orders, i.e., we plotted the histogram of

$$\eta = 10\log_{10}\epsilon_w/D^2 - 10\log_{10}\epsilon_l. \tag{20}$$

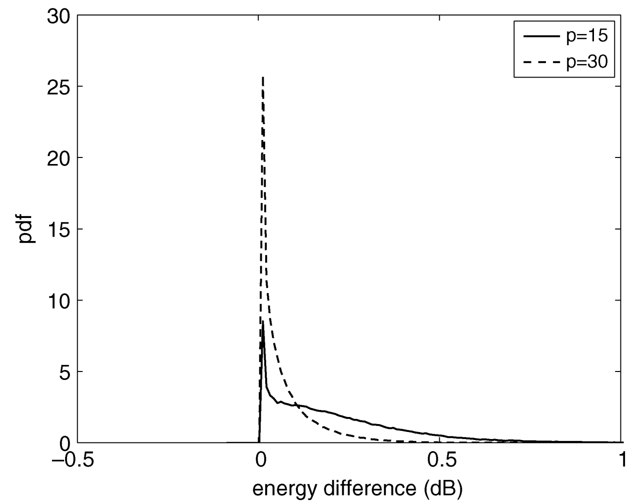

Fig. 7. Estimated probability density function (pdf) of the residual energy difference η after optimization per frame for prediction orders 15 and 30.

We observe that the distribution is zero for negative values meaning that the Laguerre system always yields a lower energy of the signal after optimization. In principle, this means that the Laguerre system is doing a better job but, to be fair, this difference is rather small. We will explain the finding that the Laguerre system always gives less energy later on. As expressed in (19), the lower output power of the Laguerre system implies a residual signal with a higher spectral flatness.

D. Spectral Differences

Additionally, we repeated the experiment shown in Fig. 6 for each frame. The difference μ between the two amplitude responses (in dB) was calculated, i.e.,

$$\mu(\theta) = 20\log_{10}\frac{\left|F_A(e^{j\theta})/D\right|}{|F_L(e^{j\theta})|} \tag{21}$$

and from this difference characteristic the standard deviation and the largest difference over the frequency axis was determined. This leads to a standard deviation and a largest difference per frame. In Fig. 8, we have plotted these data in the form of a probability density function (pdf) derived from the histogram for $p = 30$. From the pdf of the standard deviation, we see that its mean is about 0.5 dB, which is somewhat larger than that of the residual energy difference. It shows that in the amplitude transfer the differences are somewhat larger than one might expect from the energy difference as measured from the residual signal. Inspection of the results per frame indicate that the synthesis filter of the Laguerre system gives slightly more resonant peaks compared to the WLP-A case. This is also in line with the results of the measurements of the largest difference (either positive or negative) and its histogram as is also incorporated in Fig. 8. A positive value on the horizontal axis indicates that at the maximum difference, the synthesis filter of the Laguerre system has a larger amplitude than that of the WLP-A system. If, generally speaking, the Laguerre transfer functions are somewhat more peaky, one would indeed expect the mean of the estimated probability density functions to be positive. Note however that in practical settings where we would use spectral smoothing

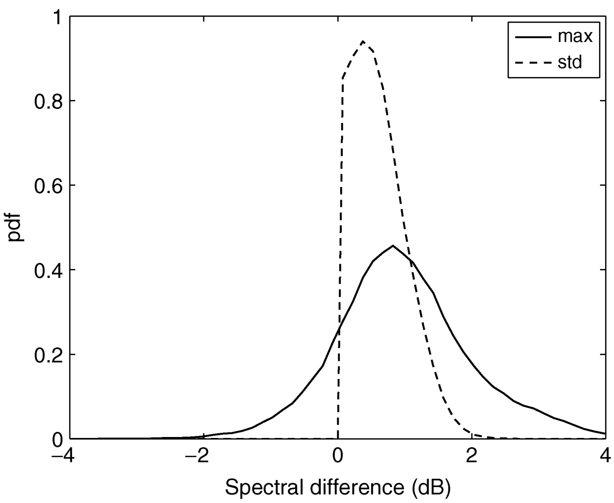

Fig. 8. Estimated pdf of largest difference and standard deviation of the spectral curves per frame for prediction order 30.

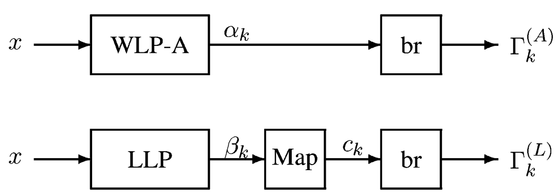

Fig. 9. Reflection coefficients associated with the WLP-A and the LLP system. The reflection coefficients of the WLP-A system are defined in the ordinary way; here shown as a mapping (br; backward recursion) from the set of polynomial coefficients α_k to the set $\Gamma_k^{(A)}$. For the LLP system, the reflection coefficients $\Gamma_k^{(L)}$ are derived via a two-stage mapping.

(bandwidth expansion) as postprocessing on the optimal coefficients, the differences will presumably become considerably less.

E. Reflection Coefficients

The fact that the spectral flatness of the Laguerre system is always larger than that of the WLP-A system seems peculiar. More surprising is the following. In [20], a mapping was proposed of the optimal Laguerre filter to a warped filter. For this mapped filter, it was experimentally found [21] that all the reflection coefficients except the last one are exactly equal to those of the optimal WLP-A solution. Obviously, this immediately explains the earlier finding that the transfer functions of these systems are so remarkably similar (Fig. 6) and the small differences as shown in Fig. 7.

The situation is depicted in Fig. 9. The reflection coefficients associated with the WLP-A scheme are called $\Gamma_k^{(A)}$, $k = 1, \cdots, p$ and can be derived from the α-coefficients using the backward recursion (br) algorithm. The reflection coefficients associated with the LLP scheme are denoted as $\Gamma_k^{(L)}$, $k = 1, \cdots, p$ and are derived from the c-coefficients which result from a mapping of the β-coefficients. The experimental finding can now be expressed as

$$\Gamma_k^{(A)} = \Gamma_k^{(L)} \text{ for } k = 1, \cdots, p-1.$$

We will prove the equivalence between the reflection coefficients associated with the two schemes, but before doing so, we will first take a closer and slightly more general look at the warped and Laguerre filters. In this way, we can explain the mapping shown in Fig. 9 and introduced in [20] for the purpose of quantization of the Laguerre prediction parameters. Furthermore, this general look reveals that the LLP is actually a WLP system with a very logical optimization criterion.

IV. Warped and Laguerre Filters

We will present some definitions which will serve us later. Note that we return to the definition of warped and Laguerre filters where the parameters of these filters are not necessarily defined by some minimization criterion.

Definition: A pth-order warped feed-forward filter is defined as a filter with transfer function

$$F_1(z) = \sum_{k=0}^{p} a_k A^k(z) \tag{22}$$

with A being a first-order allpass section as defined in (4) and a_k, $k = 0, \cdots, p$ denoting the filter coefficients.

Similarly, we define for our context the Laguerre filter as follows.

Definition: A pth-order Laguerre filter is defined as a filter with transfer function

$$F_2(z) = b_0 + \frac{z^{-1}\sqrt{1-\lambda^2}}{1-\lambda z^{-1}} \sum_{k=1}^{p} b_k A^{k-1}(z) \tag{23}$$

with A being a first-order allpass section as defined in (4) and b_k, $k = 0, \cdots, p$ denoting the filter coefficients.

The class of pth-order warped feed-forward filters is equivalent to that of the pth-order Laguerre filters. This means that given a set of coefficients a_k, we can find a set of coefficients b_k such that $F_1 = F_2$. The relation between the a_k's and b_k's is given by

$$\begin{aligned} a_0 &= b_0 + \lambda b_1/\Lambda, \\ a_i &= (b_i + \lambda b_{i+1})/\Lambda, \text{ for } i = 1, \cdots, p-1, \\ a_p &= b_p/\Lambda \end{aligned} \tag{24}$$

where $\Lambda = \sqrt{1-\lambda^2}$.

The proof is straightforward. We note that

$$\frac{z^{-1}\sqrt{1-\lambda^2}}{1-\lambda z^{-1}} = \frac{1}{\Lambda}\left[\lambda + \frac{-\lambda + z^{-1}}{1-\lambda z^{-1}}\right].$$

Therefore, (23) becomes

$$\begin{aligned} F_2(z) &= b_0 + \frac{1}{\Lambda}\sum_{k=1}^{p} b_k \left[\lambda + \frac{-\lambda + z^{-1}}{1-\lambda z^{-1}}\right] A^{k-1}(z) \\ &= b_0 + \frac{1}{\Lambda}\sum_{k=0}^{p-1} \lambda b_{k+1} A^k(z) + \frac{1}{\Lambda}\sum_{k=1}^{p} b_k A^k(z). \end{aligned}$$

From this, (24) is obvious.

We note that the definition of these functions is slightly more general than those used in the linear prediction schemes; there we use $a_0 = 1$ and $b_0 = 1$. We therefore introduce the following terminology.

Definition: A normalized pth-order warped feed-forward filter is a pth-order warped feed-forward filter with $a_0 = 1$.

The set of parameters of this normalized warped filter is associated with a monic polynomial.

Definition: A normalized pth-order Laguerre filter is a pth-order Laguerre filter with $b_0 = 1$.

We note that these two classes of normalized filters are not equivalent. When we map the normalized pth-order Laguerre filter to the pth-order warped feed-forward filter we have a linear constraint on the warped filter coefficients, namely

$$\sum_{k=0}^{p} a_k(-\lambda)^k = 1. \tag{25}$$

Conversely, when we map the normalized pth-order warped feed-forward filter to the pth-order Laguerre filter we have a linear constraint on its coefficients, namely

$$b_0 + b_1\lambda/\Lambda = 1. \tag{26}$$

From the foregoing, we infer the following. The design of a normalized pth-order Laguerre by minimization of the output error energy is equivalent to the design of a pth-order warped feed-forward filter using minimization of the output signal energy where the coefficients of the warped filter adhere to the constraint (25) instead of $a_0 = 1$.

We call this scheme WLP-C since it is clearly an alternative to the schemes WLP-A and WLP-B. The WLP-C scheme is thus defined as follows.

WLP-C An optimal warped linear predictor is defined as a warped feed-forward filter where the coefficients are optimized according to minimization of the criterion

$$E\left[\left(a_0x(n) + \sum_{k=1}^{p} a_k x_k(n)\right)^2\right]$$

under the linear constraint

$$\sum_{k=0}^{p} a_k(-\lambda)^k = 1.$$

The optimal coefficients of the filter are called γ_k, $k = 0, \cdots, p$.

The filter resulting from the optimization defined by WLP-C is identical to the optimal LLP; i.e., instead of solving the WLP-C optimization, the γ_k's can be obtained by calculating the β_k's from the LLP system and substituting these in (24) for the b_k's.

Since the result of this optimization is identical to the LLP optimization in terms of the obtained filter, we conclude that the WLP-C system has the whitening property [8], that an average spectral amplitude transfer of 0 dB is inherently incorporated in the optimization procedure and that it results in minimum-phase filters (when using the autocorrelation method). This explains why, in the experimental comparison discussed in Sections III-B and III-C, it was found that the Laguerre system always yielded a lower residual energy and a higher spectral flatness; WLP-C attains by definition the minimum output signal power of any warped feed-forward system restricted by an average 0-dB spectral amplification. In fact, the WLP-A system with rescaling in order to obtain the average spectral amplification of 0 dB is a (in practice slightly) suboptimal way of doing the same.

Note that in the conventional linear prediction case we have $a_0 = 1$ to ensure whitening; in WLP-C this constraint is changed to incorporate the whitening as a feature of the optimization. This is why we consider WLP-C/LLP as the logical extension of the conventional LP definition. It can also be observed that for $\lambda = 0$ the above constraint reduces to the conventional linear prediction constraint.

We now consider the proposed mapping in [20]. We have seen that we can map a normalized pth-order Laguerre filter to a pth-order warped feed-forward filter constrained by (25). This is a linear mapping of the coefficients. Next we can map the pth-order warped feed-forward filter with said constraint onto the normalized pth-order warped feed-forward filter by normalization of the coefficients according to

$$c_i = \gamma_i/\gamma_0, \text{ for } i = 0, \cdots, p. \tag{27}$$

Obviously, when $\gamma_0 = 0$ we would have a problem. However, in the case that the normalized pth-order Laguerre filter is a minimum-phase filter (which is the case if we design it as outlined in Section V-B), the whole mapping $\{\beta_k\} \rightarrow \{\gamma_k\} \rightarrow \{c_k\}$ forms an invertible operation [20]. The explicit expression for the c-coefficients in terms of the β_k's reads

$$c_i = \begin{cases} 1 & \text{for } i = 0, \\ \frac{\beta_i + \lambda\beta_{i+1}}{\Lambda + \lambda\beta_1}, & \text{for } i = 1, \cdots, p-1, \\ \frac{\beta_p}{\Lambda + \lambda\beta_1}, & \text{for } i = p. \end{cases} \tag{28}$$

V. Equivalence of Reflection Coefficients

In this section, we will give the normal equations for the WLP-A system (Section V-A), the LLP system (Section V-B) and finally prove the equivalence of the reflection coefficients (except the last one) of both systems (Section V-C).

A. Warped Linear Prediction (WLP-A)

As suggested by Strube in [5], the warped linear predictor filter is supplemented with a prefilter $W(z)$ to minimize the error on the nonwarped frequency axis. The input, $x(n)$, is prefiltered by $W(z)$ and the resulting signal is now denoted by $y_0(n)$. Similarly, the observed signal at the output of the kth allpass section in the filter structure is denoted by $y_k(n)$. This is illustrated in Fig. 2. The error signal, $e_w(n)$, can be written as

$$e_w(n) = y_0(n) + \sum_{k=1}^{p} a_k y_k(n). \tag{29}$$

The optimal parameters for a_k are denoted as α_k and are obtained by minimizing a mean squared-error criterion. The mean squared-error criterion, ϵ_w, is defined as

$$\epsilon_w = E\left[e_w^2(n)\right]. \tag{30}$$

Minimization of (30) leads to the optimal parameters α_k given by

$$\sum_{k=1}^{p} \alpha_k E\left[y_k(n)y_i(n)\right] = -E\left[y_0(n)y_i(n)\right] \qquad (31)$$

for $1 \leq i \leq p$. The only difference between (9) and (31) is that the observed cross-powers in (9) are now derived from signals x_k which have been subject to an additional prefiltering by $W(z)$.

With the definition of $r(i)$ according to

$$r(i) = E\left[y_i(n)y_0(n)\right] \qquad (32)$$
$$r(-i) = r(i) \qquad (33)$$

and due to the fact that

$$E\left[y_{i+k}(n)y_{j+k}(n)\right] = E\left[y_i(n)y_j(n)\right] = r(i-j) \qquad (34)$$

(assuming real-valued signals; see Appendix), we arrive at the YW equations

$$\begin{bmatrix} r(0) & \dots & r(p-1) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ r(p-1) & \dots & r(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_p \end{bmatrix} = - \begin{bmatrix} r(1) \\ \cdot \\ \cdot \\ \cdot \\ r(p) \end{bmatrix}. \qquad (35)$$

Equation (35) can be expressed in shorthand notation as

$$R_w \underline{\alpha} = -\underline{q}_w \qquad (36)$$

where R_w is the $p \times p$ Gram matrix, $\underline{\alpha} = [\alpha_1 \alpha_1 \dots \alpha_p]^T$ and $\underline{q}_w$ are $p \times 1$ vectors.

Due to the symmetry and Toeplitz structure of R_w, this system of equations can be solved in a computationally efficient manner using the Levinson–Durbin algorithm [16]. The set of reflection coefficients $\Gamma_k^{(A)}$ is obtained as a by-product during the Levinson–Durbin recursive procedure. These reflection coefficients represent the parameters associated with the lattice filter realization of the above predictor.

B. Laguerre Linear Prediction (LLP, WLP-C)

The transfer function of the LLP scheme is expressed as

$$F_L(z) = 1 + z^{-1} \sum_{k=1}^{p} b_k \frac{\sqrt{1-\lambda^2}}{1-\lambda z^{-1}} \left(\frac{-\lambda + z^{-1}}{1-\lambda z^{-1}}\right)^{k-1}. \qquad (37)$$

From Fig. 5, it can be seen that the LLP is made of the same filter sections as those used in the definition of the optimal coefficients according to the WLP-A scheme (Fig. 2) except for an additional delay element z^{-1}. For convenience, we therefore assume that if we have the sequence $x(n)$ as input to the WLP-A scheme, we have $x(n+1)$ as input to the LLP scheme. If x is a stationary stochastic signal, then both signals have the same stochastic properties. Using $x(n+1)$ as input to the LLP scheme, we have exactly the same signals y_k at the output of the allpass line in the LLP system as appeared in the top branch of the WLP-A scheme.

The optimal filter coefficients β_k are obtained by minimizing the following mean-squared error with respect to the b_k's

$$\epsilon_l = E\left[e_l^2\right] = E\left[\left\{x(n+1) + \sum_{k=1}^{p} b_k y_{k-1}(n)\right\}^2\right]. \qquad (38)$$

A set of normal equations similar to (31) can be obtained by minimizing (38). The normal equations in this case are expressed as

$$\sum_{k=1}^{p} \beta_k E\left[y_k(n)y_i(n)\right] = -E\left[x(n+1)y_i(n)\right]$$

for $0 \leq i \leq p-1$. As a shorthand notation we use

$$R_l \underline{\beta} = \underline{q}_l. \qquad (39)$$

where $\underline{\beta} = [\beta_1\ \beta_2\ \cdots\ \beta_p]^T$ and $\underline{q}_l = [q_0\ q_1\ \cdots q_{p-1}]^T$. It can be observed that R_l is identical to R_w, i.e., $R_l = R_w$. However, the elements of $\underline{q}_l$ are different from those of $\underline{q}_w$ since $q(i) = E[x(n+1)y_i(n)]$. However, the elements in $\underline{q}_w$ and $\underline{q}_l$ are not unrelated; in fact we have (see the Appendix)

$$\Lambda r(k) = q(k-1) + \lambda q(k). \qquad (40)$$

In [20], it was proposed to map the set $\{\beta_k\}$ to a set $\{c_k\}$ associated with a normalized pth order warped feed-forward filter. We have discussed this mapping $\{\beta_k\} \rightarrow \{\gamma_k\} \rightarrow \{c_k\}$ in (28) in Section IV. A consequence of this mapping is that the parameters of an LLP can consequently be quantized similar to those of an WLP which in turn can be quantized like those of a conventional tapped-delay-line (e.g., log area ratios, line spectral frequencies).

As mentioned in Section III-E, experimental observations in [21] have shown that reflection coefficients associated with the set c_k from the LLP scheme are identical to those associated with the set a_k of the WLP-A scheme *except* for the last one. We are now ready to prove this.

C. Proof of the Equivalence

In this section, we prove the equivalence of the first $p-1$ reflection coefficients associated with the optimal coefficients α_k and those associated with c_k (and thus with the LLP/WLP-C system). In view of the one-to-one relationship between reflection coefficients and the YW equations, we can translate the experimental finding on the reflection coefficients directly into YW equations to which the set $\{c_k\}$ has to adhere. This is stated in the Lemma below.

Lemma 1: The set of coefficients $\{c_k\}$ is the solution of

$$\begin{bmatrix} r(0) & r(1) & \dots & r(p-1) \\ r(1) & r(0) & \ddots & r(p-2) \\ \vdots & \ddots & \ddots & \vdots \\ r(p-1) & r(p-2) & \dots & r(0) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(p-1) \\ u_p \end{bmatrix} \tag{41}$$

for some $u_p \in \mathbb{R}$, i.e., they are the solution of YW equations corresponding to the autocorrelation sequence $\tilde{u} = [r(0), r(1), \dots, r(p-1), u_p]$.

It is clear that if this Lemma holds, then the reflection coefficients $\Gamma_k^{(A)}$ and $\Gamma_k^{(L)}$ are equal for $k = 1, \cdots, p-1$ since the kth reflection coefficient depends only on the autocorrelation sequence up to and including index k. That $\tilde{u}$ is an autocorrelation sequence stems from the fact that $\{c_k\}$ represents a minimum-phase filter [20]. We will now prove the Lemma.

Proof of Lemma 1: We show (41) by a direct computation using (28) and (40). Consider the kth equation in (41), where $k = 1, \dots, p-1$

$$\begin{aligned} &\sum_{i=1}^{p} r(k-i)c_i \\ &\quad = \sum_{i=1}^{p-1} r(k-i)\frac{\beta_i + \lambda\beta_{i+1}}{\Lambda + \lambda\beta_1} + r(k-p)\frac{\beta_p}{\Lambda + \lambda\beta_1} \\ &\quad = \frac{1}{\Lambda + \lambda\beta_1}\left[\sum_{i=1}^{p} r(k-i)\beta_i + \lambda\sum_{i=2}^{p} r(k+1-i)\beta_i\right] \\ &\quad = \frac{1}{\Lambda + \lambda\beta_1}\left[\sum_{i=1}^{p} r(k-i)\beta_i + \lambda\sum_{i=1}^{p} r(k+1-i)\beta_i\right] \\ &\qquad - \frac{\lambda r(k)\beta_1}{\Lambda + \lambda\beta_1} \\ &\quad = \frac{-q(k-1) - \lambda q(k) - \lambda r(k)\beta_1}{\Lambda + \lambda\beta_1} = -r(k). \end{aligned}$$

VI. Conclusion

We have analyzed the warped linear prediction (WLP) and Laguerre linear prediction (LLP) schemes. In order to have the whitening property in the WLP schemes, two alternatives were known. The first one (WLP-A) prefilters the signal before calculation of the optimal coefficients to achieve whitening. The second one (WLP-B) resorts to postfiltering of the output signal. We have shown that there is a third alternative to incorporate whitening (WLP-C), namely by invoking a different linear constraint [(25)] on the parameters than the standard one (i.e., $a_0 = 1$). Furthermore, we have shown that this latter procedure (WLP-C) is identical to LLP. Finally, we have shown that the optimal filter defined by the WLP-A scheme is almost identical to that of the LLP scheme: all associated reflection coefficients except the last one are identical.

Our theoretical analysis reveals that there are very tight links between the WLP and LLP system. It explains and underpins what was already known from practice: for a sufficiently high order the WLP and LLP system produce almost identical results when considering, e.g., their transfer characteristics. We now can more firmly state that experimental results (e.g., the performance in terms of flattening, quality in coding, parameter bit rate) of one of the cases (i.e., WLP-A, WLP-B, WLP-C/LLP) will carry over to the other cases. The difference between the cases for prediction orders used in practice is more in the implementation. The structure associated with LLP system lends itself immediately for an identical implementation of the predictor in the analysis and synthesis filter. This guarantees perfect reconstruction even in case of finite word-length arithmetic.

All of the experimental and theoretical results were obtained for output power minimization. As shown in Section IV, the subspaces associated with a pth order WLP-A and LLP systems are nearly identical and thus the information contained in the regressor signals is almost identical. Therefore, we argue that the main conclusion will remain the same for other optimization criteria, i.e., the WLP and LLP systems produce almost identical results for sufficiently high order and in that case the actual difference is more of an implementation issue.

Appendix

Here, we give the straightforward proofs of symmetric and Toeplitz character of the matrix R in (10), of (40), and of (19). The proof of (34) is completely analogous to that regarding the properties of the entries of R and is therefore omitted. We assume real-valued, wide-sense stationary signals. All integrals are taken over the interval $[0, 2\pi)$, $z = e^{j\theta}$, and $P_x(\theta)$ is the power spectral density function of x.

We start with the properties of R. For the entries of R we have

$$\begin{aligned} E\left[x_{i+k}(n)x_{j+k}(n)\right] &= \frac{1}{2\pi}\int A^{i+k}(z)A^{j+k}(z^{-1})P_x(\theta)d\theta \\ &= \frac{1}{2\pi}\int A^{i}(z)A^{j}(z^{-1})P_x(\theta)d\theta \\ &= E\left[x_i(n)x_j(n)\right] \end{aligned}$$

from which immediately follows the symmetric and Toeplitz character of R.

Next, we prove (40) as follows:

$$\begin{aligned} &q(k-1) + \lambda q(k) \\ &\quad = E\left[x(n+1)\left\{y_{k-1}(n) + \lambda y_k(n)\right\}\right] \\ &\quad = \frac{1}{2\pi}\int \left[zP_x(\theta)W(z^{-1})A^{k-1}(z^{-1})\right. \\ &\qquad\qquad \left. + \lambda zP_x(\theta)W(z^{-1})A^{k}(z^{-1})\right] d\theta \\ &\quad = \frac{1}{2\pi}\int z\left[A^{-1}(z^{-1}) + \lambda\right] P_x(\theta)W(z^{-1})A^{k}(z^{-1})d\theta \\ &\quad = \frac{1}{2\pi}\int \left[\frac{1-\lambda^2}{1-\lambda z^{-1}}\right] P_x(\theta)W(z^{-1})A^{k}(z^{-1})d\theta \\ &\quad = \frac{1}{2\pi}\int \Lambda W(z)P_x(\theta)W(z^{-1})A^{k}(z^{-1})d\theta \\ &\quad = \Lambda E\left[y_0(n)y_k(n)\right] = \Lambda r(k). \end{aligned}$$

The proof (19) follows from the following equalities. First, we have the equivalence $\gamma_w = \gamma'_w$ from the definition (18)

$$(\gamma'_w)^2 = \frac{\exp\left\{\frac{1}{2\pi}\int \ln P_{e'_w}(\theta)d\theta\right\}}{\frac{1}{2\pi}\int P_{e'_w}(\theta)d\theta} = \frac{\exp\left\{\frac{1}{2\pi}\int \ln\left[P_{e_w}(\theta)/D^2\right]d\theta\right\}}{\frac{1}{2\pi}\int P_{e_w}(\theta)/D^2 d\theta} = \gamma_w^2.$$

Thus, we have $\gamma_l/\gamma_w = \gamma_l/\gamma'_w$. Next, we have

$$\exp\left\{\frac{1}{2\pi}\int \ln P_{e'_w}(\theta)d\theta\right\} = \exp\left\{\frac{1}{2\pi}\int \ln P_x(\theta)d\theta\right\} = \exp\left\{\frac{1}{2\pi}\int \ln P_{e_l}(\theta)d\theta\right\}$$

since the averages of the logarithm of the amplitude transfers of H_A/D and H_L are 0. Lastly, $\epsilon_l = E[e_l^2] = (1/2\pi)\int P_{e_l}(\theta)d\theta$ and, similarly, $\epsilon'_w = (1/2\pi)\int P_{e_w}(\theta)d\theta/D^2$.

REFERENCES

[1] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*. Berlin, Germany: Springer-Verlag, 1976.

[2] J. Makhoul, "Linear prediction: A tutorial review," *Proc. IEEE*, vol. 63, no. 4, pp. 561–579, Apr. 1975.

[3] P. P. Vaidyanathan, *The Theory of Linear Prediction*. San Rafael, CA: Morgan & Claypool, 2007.

[4] A. Oppenheim, D. Johnson, and K. Steiglitz, "Computation of spectra with unequal resolution using the fast Fourier transform," *Proc. IEEE*, vol. 59, no. 2, pp. 299–301, Feb. 1971.

[5] H. W. Strube, "Linear prediction on a warped frequency scale," *J. Acoust. Soc. Amer.*, vol. 68, no. 4, pp. 1071–1076, 1980.

[6] A. Harma and U. Laine, "A comparison of warped and conventional linear prediction coding," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 4, pp. 579–588, Jul. 2001.

[7] B. Edler and G. Schuller, "Audio coding using a psycho-acoustic pre- and post-filter," in *Proc. ICASSP'00*, Istanbul, Turkey, Jun. 5–9, 2000, pp. 1881–1884.

[8] A. C. den Brinker, V. Voitishchuk, and S. J. L. van Eijndhoven, "IIR-based pure linear prediction," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 1, pp. 68–75, Jan. 2004.

[9] A. Biswas and A. C. den Brinker, "Perceptually biased linear prediction," *J. Audio Eng. Soc.*, vol. 54, pp. 1179–1188, 2006.

[10] A. C. den Brinker, J. Breebaart, P. Ekstrand, J. Engdegård, F. Henn, K. Kjörling, W. Oomen, and H. Purnhagen, "An overview of the coding standard MPEG-4 Audio Amendments 1 and 2: HE-AAC, SSC and HE-AAC v2," *EURASIP J. Audio, Speech, Music Process.*, vol. 2009, 2009, Article ID 468971.

[11] A. C. den Brinker and A. Biswas, "Quantization of Laguerre-based stereo linear predictors," presented at the Proc. 122nd AES Conv., Vienna, Austria, May 2007, Conv. Paper 7006.

[12] J. O. Smith and J. S. Abel, "Bark and ERB bilinear transforms," *IEEE Trans. Speech Audio Process.*, vol. 7, no. 6, pp. 697–708, Nov. 1999.

[13] A. El-Jaroudi and M. J. , "Discrete all-pole modeling," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 441–423, Feb. 1991.

[14] M. N. Murthi and B. D. Rao, "All-pole modelling of speech based on the minimum variance distortionless response spectrum," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 221–239, May 2000.

[15] C. Magi, T. Bäckström, and P. Alku, "Objective and subjective evaluation of seven selected all-pole modelling methods," in *Proc. 7th Nordic Signal Process. Symp.*, Rejkjavik Iceland, Jun. 7–9, 2006, pp. 118–121.

[16] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: Wiley, 1996.

[17] A. Härmä, "Implementation of frequency-warped filters," *Signal Process.*, vol. 80, pp. 543–548, 2000.

[18] P. W. Broome, "Discrete orthonormal sequences," *J. Assoc. Comput. Mach.*, vol. 12, no. 2, pp. 151–165, Apr. 1965.

[19] T. Oliveira e Silva, "Laguerre filters—An introduction," *Revista Do Detua*, vol. 1, no. 3, pp. 237–248, Jan. 1995.

[20] A. C. den Brinker and F. Riera-Palou, "Pure linear prediction," presented at the Proc. 115th AES Conv., New York, Apr. 18, 2003, Conv. Paper 5924.

[21] A. Biswas, "Advances in perceptual stereo audio coding using linear prediction techniques," Ph.D. dissertation, Technical Univ. Eindhoven (TUE), Eindhoven, The Netherlands, May 2007.

Albertus C. den Brinker (M'03–SM'07) received the M.Sc. degree in electrical engineering and the Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1983 and 1989, respectively.

From 1987 to 1999, he worked in the Signal Processing Group, Faculty of Electrical Engineering, Eindhoven University of Technology. In 1999, he joined the Digital Signal Processing Group at Philips Research Laboratories, Eindhoven, where he is head of the Signal Processing Techniques for Audio and Speech cluster. One of the activities within the cluster concerns standardization of audio coders, especially standardization within MPEG. Major contributions were made to MPEG-4 Amendment 2 (high-quality parametric audio coding, also known as MPEG-4 SSC) and MPEG Surround. He publishes regularly in international scientific journals and proceedings of scientific conferences and is author and coauthor of several patents.

Harish Krishnamoorthi (S'10) was born in Chennai, India, in 1984. He received the B.Sc. degree in electronics and communication engineering from PSG College of Technology, Coimbatore, India, in 2005 and the M.S. degree in electrical engineering from Arizona State University (ASU), Tempe, in 2007. He is currently pursuing the Ph.D. degree in electrical engineering at ASU.

From May 2008 to August 2008, he was a Visiting Researcher in the Signal Processing Group, Eindhoven University of Technology, Eindhoven, The Netherlands. His primary research interests are in the area of speech and audio coding, speech enhancement, and psychoacoustics.

Evgeny A. Verbitskiy received M.Sc. degrees in mathematics from the University of Groningen, Groningen, The Netherlands, in 1996 and Moscow State University, Moscow, Russia, in 1997 and the Ph.D. in mathematics from University of Groningen in 2000.

His research interests include dynamical systems and probability theory. He was with Philips Research, Eindhoven, The Netherlands, from 2002 until 2010. Since 2007, he has held a part-time appointment as a Professor of mathematics of life sciences at the University of Groningen.