Linear Predictive Method for Improved Spectral Modeling of Lower Frequencies of Speech with Small Prediction Orders

Paavo Alku and Tom Bäckström

Abstract— An all-pole modeling technique, Linear Prediction with Low-frequency Emphasis (LPLE), which emphasizes the lower frequency range of the input signal, is presented. The method is based on first interpreting conventional linear predictive (LP) analyses of successive prediction orders with parallel structures using the concept of symmetric linear prediction. In these implementations, symmetric linear prediction is preceded by simple pre-filters, which are of either low or high frequency characteristics. Combining those symmetric linear predictors that are not preceded by high-frequency pre-filters yields the proposed LPLE predictor. It is proved that the all-pole filters computed by LPLE are always stable. The results achieved with vowels show that the proposed method is well-suited for those applications, where low-order all-pole models with improved modeling of the lowest formants, are needed.

Index Terms— all-pole modeling, linear prediction, LSP decomposition

I. INTRODUCTION

Linear prediction (LP) is among the most widely used methods of speech processing. Especially in low bit rate speech coding, LP has an established role and many of the standardized coders are based on linear prediction [1], [2], [3]. Among the different variations of LP, the classical autocorrelation method of linear prediction is the most popular. In this method, a predictor (an FIR of order m) is determined by minimizing the square of the prediction error, i.e. the residual, over a time interval, which is, in principle, infinite [4]. The prevalence of the conventional autocorrelation method of LP is explained by the fact that it yields stable all-pole models for speech spectra, which are accurate enough for most applications when defined by a few parameters. Moreover, computation of the filter parameters can be done by applying fast algorithms such as the Levinson-Durbin method [4].

P. Alku and T. Bäckström are with the Helsinki University of Technology (HUT), Laboratory of Acoustics and Audio Signal Processing, P.O.Box 3000, FIN-02015 HUT, Finland, E-mail:paavo.alku@hut.fi
EDICS: CODI ANLS

In order to model the envelope of speech spectrum accurately enough with LP, the prediction order is typically adjusted to equal the sampling frequency in kHz added by a small integer [5]. This choice of the prediction order makes it possible to compute all-pole filters that are capable to model, on average, one formant per kilohertz plus the overall spectral structure of speech caused by the glottal excitation and the lip-radiation effect. Hence, application of LP in coding of telephone band speech (i.e., bandwidth between 300 Hz and 3400 Hz, sampling frequency 8 kHz), for example, is typically based on the prediction order $m = 10$ e.g., [2], [3]. If one attempts to compress the LP information by using prediction orders smaller than those given by the rule mentioned above, the formant structure of speech becomes typically poorly modeled. This effect is demonstrated in Fig. 1, which shows an LP spectrum computed for a vowel (vowel /u/, sampling frequency 8000 kHz) using a prediction order $m = 8$, which is slightly too small. The figure illustrates, how the all-pole spectrum given by the conventional LP analysis does not separate properly the lowest two formants.

Since the criterion of optimization in the conventional LP analysis is the minimization of the residual energy, all the frequencies of the input signal are treated equally. In other words, the all-pole model computed by the conventional LP favors high-energy regions of signal spectrum no matter at which frequencies these occur. This equal treatment of the frequencies of the input signal is inconsistent with, for example, the properties of human hearing, which is known to be frequency dependent (i.e, the spectral resolution decreases towards higher frequencies) [6]. Therefore, linear predictive methods have been developed that utilize the frequency selectivity of human hearing [7], [8] [9] [10]. The equal treating of the input frequencies embedded in the conventional LP analysis is also inconsistent from the point of view of speech production, because the most important formants, the first (F1) and second (F2) formant, are typically located at frequencies below 2 kHz. Hence, from the point of view of both speech production and perception,
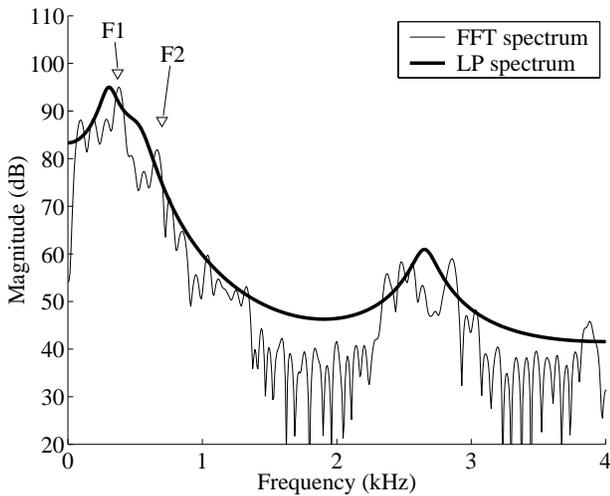
Fig. 1

ILLUSTRATION OF THE FAILURE OF CONVENTIONAL LP IN MODELING OF THE FIRST AND SECOND FORMANT (DENOTED BY F1 AND F2, RESPECTIVELY), WHEN THE PREDICTION ORDER ($m = 8$) IS SLIGHTLY TOO SMALL (THE SAMPLING FREQUENCY WAS 8KHZ). THE FFT SPECTRUM OF THE SIGNAL (VOWEL /U/) IS DEPICTED BY THIN LINE AND THE ALL-POLE SPECTRUM GIVEN BY CONVENTIONAL LP IS DEPICTED BY THICK LINE.

it would be desirable to obtain all-pole models of speech with improved resolution on the frequency range where the two lowest formants are located rather than modeling high-energy regions over the entire frequency range of the input signal.

The rationale for the present study is to modify the conventional LP analysis in order to improve modeling of the main spectral features of speech at low frequencies with small prediction orders. We will present a new linear predictive method, Linear Prediction with Low-frequency Emphasis (LPLE), which gives stable all-pole models, which, in turn, put more emphasis on the lowest frequencies of the input signal. As background for the proposed method, we will first present in Section II, how conventional LP analysis can be interpreted with the help of the Line Spectrum Pair (LSP) decomposition and symmetric linear prediction [11], [12], [13]. Using this interpretation, it is then possible to combine symmetric polynomials embedded in conventional LP analyses of successive prediction orders to define the proposed all-pole filters. The stability of the filters given by LPLE can be easily proved as shown at the end of Section III. Finally, the performance of the new method is compared to that of conventional LP by analysing the vowels of the Finnish language.

II. BACKGROUND

In this section, we will briefly present three issues, which form the basis of the proposed new algorithm: conventional linear prediction, the Line Spectral Pairs decomposition, and symmetric linear prediction.

A. Conventional linear prediction

Conventional linear prediction with the prediction order equal to m can be presented in matrix notation as follows [4]. Given the signal $x(n)$ and the predictor parameters a_i ($0 \leq i \leq m$), denoted by $\mathbf{x} = [x(n) \ldots x(n-m)]^T$ and $\mathbf{a} = [a_0 \ldots a_m]^T$ in vector notation, respectively, we can express the residual as $e(n) = \mathbf{x}^T \mathbf{a}$. The optimal predictor is defined by minimizing the expected value of the residual energy $E[e^2(n)] = E[\mathbf{a}^T \mathbf{x}\mathbf{x}^T \mathbf{a}] = \mathbf{a}^T \mathbf{R}\mathbf{a}$ subject to the constraint $a_0 = 1$ or equivalently $\mathbf{a}^T \mathbf{b} = 1$, where $\mathbf{b} = [1\, 0 \ldots 0]^T$. The optimal solution can be written in the form of extended normal equations as:

$$\mathbf{R}\mathbf{a} = \sigma^2 \mathbf{b} \qquad (1)$$

where σ^2 denotes the energy of the residual.

In the following section, we will denote the transfer function of the optimal predictor given by conventional LP (the Z-transform of sequence $\mathbf{a}$ defined in Eq. 1) by $A(z)$.

B. Line Spectral Pair (LSP) decomposition

Given an LP predictor $A(z)$ with a prediction order equal to m, the Line Spectral Pair (LSP) decomposition defines two polynomials of order $m + 1$, the symmetric polynomial $P(z)$ and the anti-symmetric polynomial $Q(z)$, as follows [11]:

$$\begin{array}{rcl} P(z) & = & A(z) + z^{-m-1}A(z^{-1}) \\ Q(z) & = & A(z) - z^{-m-1}A(z^{-1}). \end{array} \qquad (2)$$

It has been proved that the LSP-decomposition has the following essential properties [11], [14], [15], [16]:

1) The zeros of $P(z)$ and $Q(z)$ are always on the unit circle.
2) When m is even, $P(z)$ has a trivial root located at $z = -1$ and $Q(z)$ has a trivial root located at $z = 1$. When m is odd, $Q(z)$ has two trivial roots, $z = 1$ and $z = -1$, while $P(z)$ has no trivial roots.
3) When $A(z)$ is minimum-phase, zeros of $P(z)$ and $Q(z)$ are interlaced. This property is called the intramodel interlacing theorem.
4) The roots of $P(z)$ computed from an LP-predictor of order m are interlaced with the roots of $P(z)$ computed from an LP-predictor of order $m - 1$.

Similarly, roots of $Q(z)$ computed from a m'th order LP-predictor interlace with those defined from an LP-predictor of order $m-1$. This property is called the intermodel interlacing theorem.

Predictor $A(z)$ can be obtained from the LSP polynomials simply as:

$$A(z) = \frac{1}{2}\left[P(z) + Q(z)\right]. \qquad (3)$$

The LSP decomposition is a widely used method to represent LP parameters in speech coding because it is well suited for quantization [17] and interpolation [18]. The intramodel interlacing theorem is especially highly appreciated in speech coding applications because it can be used to guarantee that the all-pole computed by LP is stable after quantization. Even though the most important application area of the LSP decomposition is undoubtedly in speech coding, the method has also been used, for example, in speech recognition [19], [20] and in speaker recognition [21].

C. Symmetric linear prediction

Symmetric linear prediction [12], [13] with prediction order equal to p is based on the predictor polynomial defined as:

$$B(z) = 1 + b_1 z^{-1} + \ldots + b_{\frac{p}{2}-1} z^{-\frac{p}{2}+1} + b_{\frac{p}{2}} z^{-\frac{p}{2}} \quad (4)$$
$$+ b_{\frac{p}{2}-1} z^{-\frac{p}{2}-1} + \ldots + b_1 z^{-p+1} + z^{-p}.$$

In this structure, a predictor of order p can be defined from $p/2$ coefficients b_i $(1 \le i \le p/2)$ because of the symmetry of the impulse response. Using a similar optimization procedure as in the case of conventional LP, it can be shown that the optimal predictor for symmetric linear prediction is obtained from the following normal equations:

$$\sum_{k=1}^{\frac{p}{2}-1} b_k \left[R(k-j) + R(k-p+j)\right] + b_{\frac{p}{2}} R(j - \frac{p}{2})$$
$$= -R(j) - R(j-p), \qquad j \in \left[1, \frac{p}{2}\right], \quad (5)$$

where autocorrelation is estimated from samples of signal $x(n)$, $0 \le n \le L-1$, as $R(i) = \sum_{n=0}^{L-i-1} x(n)x(n+i)$.

It has been established that roots of $B(z)$ are always on the unit circle [12]. Essentially, there is a close connection (although perhaps not well known in the speech coding community) between symmetric LP and the LSP decomposition: the LSP polynomials (excepting the points $z = \pm 1$) are, in fact, LP predictors, which minimize the energy of the prediction error subject to the

constraint that the zeros of the predictor are restricted to the unit circle, i.e., the residual is computed using symmetric linear prediction [13].

Using the relationship between the LSP decomposition and the symmetric linear prediction, we can now express different implementations for the conventional linear prediction as shown in the flow graphs in Fig. 2. The figure depicts different alternatives for the implementation of the conventional linear prediction regardless of whether the prediction order is even or odd. Normally, conventional linear prediction is computed by inverse filtering $x(n)$ using the optimal predictor $A(z)$ defined in Eq. 1 (Fig. 2(a)). With the LSP decomposition (Eq. 3), however, it is possible to implement the same filtering using the parallel structure shown in Fig. 2(b). This implementation requires more computations than the basic alternative shown in Fig. 2(a) and is therefore useless as such. However, it serves as an intermediate step to the implementations shown in Fig. 2(c) and 2(d), where conventional linear prediction is also implemented as a parallel structure. In this alternative, the trivial roots of the LSP polynomials have been separated and, consequently, both $P(z)$ and $Q(z)$ can be expressed according to [13] with the help of symmetric LP predictors. For even values of the prediction order (see Section II-B), there is a single trivial root to be separated from both $P(z)$ and $Q(z)$ (Fig. 2(c)). For odd values of m, there are two trivial roots to be separated only from $Q(z)$ (Fig. 2(d)). Hence, the transfer function of the conventional LP predictor computed by the structure shown in Fig. 2(c) can be expressed using the symmetric linear prediction for even values of the prediction order as follows:

$$A(z) = 1/2\left[B_{1,e}(z)P_1(z) + B_{2,e}(z)P_2(z)\right], \quad (6)$$

where $B_{1,e}(z)$ and $B_{2,e}(z)$ are transfer functions of the symmetric linear predictors (both of the order m) computed from $x(n)$ filtered through pre-filters $P_1(z) = 1 + z^{-1}$ and $P_2(z) = 1 - z^{-1}$, respectively. For odd values of m, the transfer function of the conventional LP analysis can be expressed as:

$$A(z) = 1/2\left[B_{1,o}(z) + B_{2,o}(z)P_1(z)P_2(z)\right], \quad (7)$$

where $B_{1,o}(z)$ is the symmetric linear predictor (of order $m + 1$) computed directly from $x(n)$ and $B_{2,o}(z)$ is the symmetric linear predictor (of order $m - 1$) computed from $x(n)$ filtered through a cascade of filters $P_1(z) = 1 + z^{-1}$ and $P_2(z) = 1 - z^{-1}$.

Implementation of conventional LP using the structure shown in Fig. 2(c) and 2(d) is, again, inferior to the one shown in Fig. 2(a) in terms of the computational load. However, this interpretation of conventional LP serves as

a starting point for the new all-pole modeling technique described next in Section III-A.

III. METHOD

A. LPLE Algorithm

Similarly to the example shown in Fig. 1, assume that a given vowel sound sampled with 8 kHz and conventional LP analysis of an even prediction order is to be computed. (It is germane to consider only even values of m since modeling one spectral resonance requires at least one complex conjugate pair of poles. Hence, modeling a maximum number of formants with an all-pole filter of order m calls for using an even prediction order. Consequently, a vast majority of LP applications use an even value of the prediction order.) Assume also that the prediction order is smaller than that required by the sampling frequency (e.g., $m = 8$). According to the flow graph shown in Fig. 2(c), this conventional LP predictor corresponds to defining two symmetric linear predictors, both of which are defined by filtering the input signal through a simple, fixed pre-filter. It is worth noticing in this implementation that the zeros of $P_1(z)$ and $P_2(z)$ are located at $z = -1$ and $z = +1$, respectively, which implies that the amplitude response of these two pre-filters at the low and the high end of the frequency range is considerably different (Fig. 3). Therefore, in determining the LP predictor according to Fig. 2(c), the symmetric linear predictor in the upper branch will most likely not locate any of its roots at the high end of the frequency band due to extensive damping of these frequencies by $P_1(z)$. (Recall that symmetric LP is based on the mean square error criterion, which means that it focuses on the strongest spectral components.) Similarly, roots of $B_{2,e}(z)$ are not likely to occur at the lowest part of the frequency range due to extensive attenuation of these frequencies by $P_2(z)$. If there is a great distance between a root of $B_{1,e}(z)$ (on the unit circle) and its counterpart of $B_{2,e}(z)$ (also on the unit circle), the corresponding root of the final LP predictor (located inside the unit circle) will be at a distance from the unit circle. (This follows from properties between LSP roots and LP, e.g. [20]). Hence, the spectral model given by LP will most likely not show a strong resonance in the frequency range in question. This phenomenon is most severe when the prediction order is too small as compared to the bandwidth of the input signal and it therefore explains the poor modeling of the lowest formants in Fig. 1 as well.

In order to obtain an all-pole filter, which puts more emphasis on the lower frequency range of the input signal, it is possible to combine the structures shown
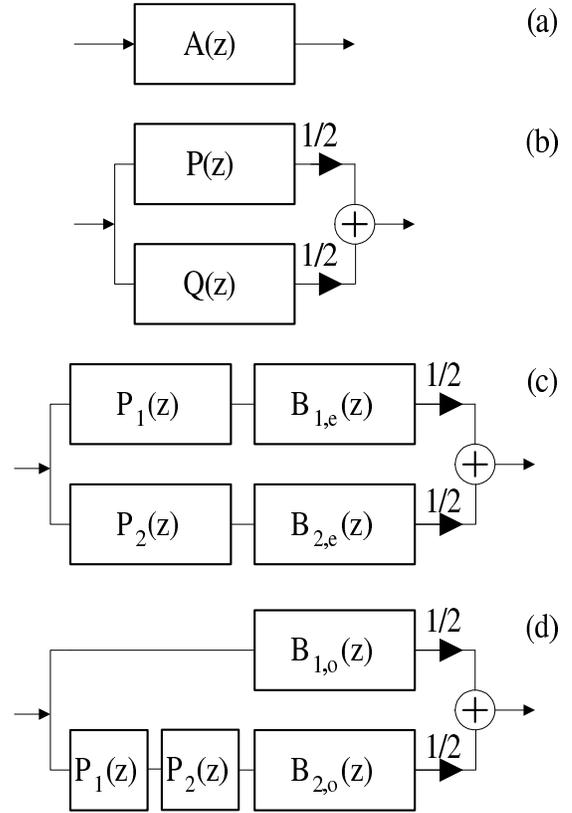
Fig. 2

DIFFERENT IMPLEMENTATIONS OF CONVENTIONAL LINEAR PREDICTION WITH PREDICTION ORDER m. (A) CONVENTIONAL IMPLEMENTATION WITH PREDICTOR $A(z)$ (EQ. 1). (B) IMPLEMENTATION BASED ON THE LSP DECOMPOSITION (EQ. 3). (C) IMPLEMENTATION WITH SYMMETRIC LP (BASED ON [12]), m EVEN. SYMMETRIC LINEAR PREDICTORS $B_{1,e}(z)$ AND $B_{2,e}(z)$ ARE OPTIMIZED USING EQ. 5 WITH $p = m$. AUTOCORRELATIONS REQUIRED IN DETERMINING $B_{1,e}(z)$ AND $B_{2,e}(z)$ ARE COMPUTED FROM $x(n)$ FILTERED THROUGH FILTERS $P_1(z) = 1 + z^{-1}$ AND $P_2(z) = 1 - z^{-1}$, RESPECTIVELY. (D) IMPLEMENTATION WITH SYMMETRIC LP (BASED ON [12]), m ODD. SYMMETRIC LINEAR PREDICTOR $B_{1,o}(z)$ IS OPTIMIZED USING EQ. 5 WITH $p = m + 1$, WHERE AUTOCORRELATIONS ARE COMPUTED FROM SIGNAL $x(n)$. $B_{2,o}(z)$ IS OPTIMIZED USING EQ. 5 WITH $p = m - 1$ BY COMPUTING AUTOCORRELATIONS FROM SIGNAL $x(n)$ FILTERED THROUGH THE CASCADE OF $P_1(z) = 1 + z^{-1}$ AND $P_2(z) = 1 - z^{-1}$.
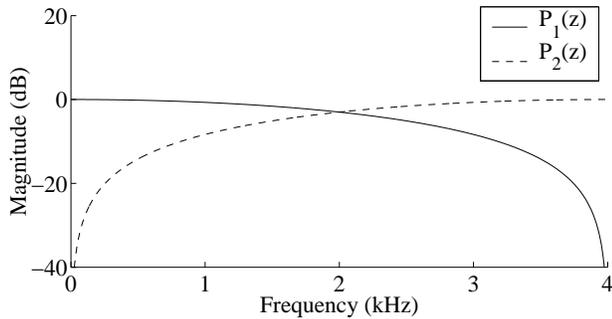
Fig. 3

AMPLITUDE RESPONSES OF PRE-FILTERS $P_1(z)$ AND $P_2(z)$ USED IN FIG. 2(C), SAMPLING FREQUENCY EQUALS 8 KHZ. MAXIMUM VALUES OF BOTH SPECTRA NORMALIZED TO 0 DB.

in Fig. 2(c) and Fix. 2(d). More specifically, we aim to take advantage of the parallel structures based on the symmetric linear prediction so as to combine the low-pass characteristics of two conventional LP analyses of *successive* prediction orders. The use of linear predictions of successive orders is motivated by the intermodel interlacing property of the LSP decomposition [15] [16], using which it is possible to easily guarantee the stability of the all-pole filter to be designed (see Section III-B). Hence, let us start by linear predictions of orders m and $m-1$ (with m even) implemented using the symmetric linear prediction shown in Fig. 2(c) and 2(d), respectively. To design an m'th order all-pole filter that emphasizes low frequencies can now be attained by combining the upper branch of Fig. 2(c), i.e., the branch with a symmetric linear prediction preceded by a fixed *low-pass filter*, to the upper branch of Fig. 2(d), i.e., the branch with a symmetric linear prediction preceded by *no pre-filter*. In order to guarantee the stability of the final all-pole filter, the symmetric linear predictor corresponding to the upper branch of Fig. 2(d) needs to be combined with a first order FIR with its zero at $z = 1$. These stages together make up the new method LPLE, the computation of which can be summarized (with prediction order equal to m) as follows:

1) From implementation of linear prediction of order m (upper branch of Fig. 2(c)): Compute the symmetric linear predictor $B_{1,e}(z)$ using Eq. 4 and 5 with $p = m$ by defining autocorrelations from $x(n)$ filtered through $P(z) = 1 + z^{-1}$.

2) From implementation of linear prediction of order $m-1$ (upper branch of Fig. 2(d)): Compute the symmetric linear predictor $B_{1,o}(z)$ using Eq. 4 and 5 with $p = (m-1) + 1 = m$ by defining autocorrelations directly from signal $x(n)$. (Notice

from Fig. 2(b) and Fig. 2(d) that $B_{1,o}(z)$ is the polynomial $P(z)$ of the LSP decomposition defined for the $(m-1)$'th order LP predictor. Hence, the order of $P(z)$, and thereby also the order of $B_{1,o}(z)$, must be one larger than $(m-1)$, i.e., the computation of $B_{1,o}(z)$ uses $p = (m-1)+1 = m$.)

3) From the two symmetric linear predictors obtained, construct the following polynomials: $S_e(z) = B_{1,e}(z)(1 + z^{-1})$ and $S_o(z) = B_{1,o}(z)(1 - z^{-1})$. Notice that $S_e(z)$ is symmetric and $S_o(z)$ is antisymmetric.

4) The final transfer function of the LPLE predictor is obtained by adding polynomials $S_e(z)$ and $S_o(z)$ and by scaling the sum by $1/2$:

$$X(z) = \frac{1}{2} \left[S_e(z) + S_o(z) \right]. \tag{8}$$

It is worth noticing that the final predictor $X(z)$ obtained by the LPLE method is of order m, but, because it is computed using LP analysis of successive orders, it is naturally sub-optimal in terms of the residual energy. In other words, there is a loss in prediction gain when the conventional LP predictor of order m is replaced with the LPLE predictor of the same order [5].

B. Stability of LPLE

The m'th order all-pole filter given in Eq. 8 can be easily proved to be stable by using the intermodel interlacing property of the LSP decomposition (property No. 4 in Section II-B) [15], [16]. According to this theory, the roots of two symmetric LSP polynomials defined for LP predictors of successive prediction orders are interlaced on the unit circle. Referring to stages No. 1 and No. 2 at the end of Section III-A, it is implied that roots of $B_{1,e}(z)$ and $B_{1,o}(z)$ must interlace because $B_{1,e}(z)$ and $B_{1,o}(z)$ are equal to the symmetric LSP-polynomials (i.e., $P(z)$) defined from LP-predictors of orders m and $m-1$, respectively, but with the trivial root at $z = -1$ of $P(z)$ computed from the m'th order LP predictor excluded. Hence, polynomials $S_e(z)$ and $S_o(z)$ in Eq. 8 have all their m complex roots interlaced on the unit circle. In addition, $S_e(z)$ and $S_o(z)$ have a single real root at $z = -1$ and $z = +1$, respectively, due to stage No. 3 in the LPLE procedure. Hence, $S_e(z)$ and $S_o(z)$ not only have an equal number of roots on the unit circle but they also interlace. This implies, according to the intramodel interlacing property, i.e, property No. 3 in Section II-B, that the sum of $S_e(z)$ and $S_o(z)$ must be minimum-phase and, consequently, the corresponding all-pole filter is always stable.

Note that the intramodel interlacing property, which was utilized in the proof above, is typically used in

literature related to the LSP polynomials computed from the m'th order predictor given by the conventional LP analysis, i.e., Eq. 2. However, the intramodel interlacing theorem holds true for arbitrary polynomials, whose roots are interlaced on the unit circle [11]. Therefore, it can be used also to prove the minimum-phase property of the sum of $S_e(x)$ and $S_o(z)$.

IV. RESULTS

A. Speech material

In order to compare the performances of conventional LP and LPLE, we recorded the eight vowels of the Finnish language produced by five female and five male speakers. Each vowel was pronounced using sustained phonation and the natural fundamental frequency of the speaker. The recording was done in an anechoic chamber using a high-quality condenser microphone (Brüel&Kjær 4188). The data were first saved onto a DAT from which the vowels were transferred into a computer using a sampling frequency of 22.050 kHz and a resolution of 16 bits. In the computer, the sounds were high-pass filtered in order to remove low-frequency fluctuations picked up in the recordings. The signals were then down-sampled to the sampling frequency of 8 kHz. The final analysis bandwidth of the vowels was between 50 Hz and 4.0 kHz.

B. Experiments

Conventional LP analysis and LPLE were computed with the prediction order $m = 8$ and the frame size of 200 samples (= 25 ms). In computing the autocorrelation terms, Hamming windowing was used. Since all-pole modeling in the experiments was deliberately computed with a prediction order that is slightly too small with respect to the sampling frequency in question, we used a conventional technique, the pre-emphasis [5], to improve the performance of low-order all-pole filters in spectral modeling. The goal of the pre-emphasis is to decrease spectral dynamics of vowel sounds prior to all-pole modeling and thereby improve modeling of higher formants. As a pre-emphasizer we used a first order FIR with its zero at $z = 0.95$. Hence, the all-pole spectra to be analyzed were obtained by cascading the 8th-order all-pole filters given by the two predictive methods with the de-emphasizer, i.e., the inverse of the pre-emphasis FIR. Furthermore, all-pole filters computed by LPLE yield sometimes poles located close to the unit circle especially in modeling F1. In order to alleviate the underestimation of formant bandwidths, the predictor

polynomial given by LPLE was windowed by an exponential function $w(n) = 0.98^n$, $0 \le n \le m$. It is well-known [22] that this windowing corresponds to replacing the original roots $z = z_i$, $1 \le i \le m$ of the predictor polynomial with $\hat{z} = 0.98z_i$, $1 \le i \le m$.

The two all-pole modeling techniques were compared in two experiments. In the first experiment (Experiment I), we simply extracted the number of formants indicated by the eighth order all-pole spectra. The formant was identified as a local maximum in the amplitude spectrum, i.e., the frequency at which the derivative of the all-pole amplitude spectrum changes its sign from positive to negative. In the second experiment (Experiment II), we made a small subjective test as follows. Firstly, eighth order all-pole filters were determined by the conventional LP and by LPLE for vowels /a/, /i/ and /u/ from voices produced by a male subject. (These three phonemes were selected because they are known to represent the corners of the F1-F2 representation of the vowel sounds [23].) The all-pole filters obtained were then used to synthesize sounds (duration 300 ms) by using impulse train excitations determined separately for each vowel. The position of impulses in the excitation waveform was computed with the help of a residual signal given by a higher order ($m = 16$) LP analysis: the impulse positions of the excitation were set to coincide with the main peaks of each period of the residual given by the 16th order LP analysis. This procedure ensured that the synthetic vowels were of the same pitch as the corresponding original vowel. The signals were then played to a panel of seven naive listeners. Each listener was asked to assess which one of the two synthetic candidates was perceptually closer to the original vowel. A neutral opinion stating that the panelist did not hear any difference between the two candidates, was also permitted. The listeners were allowed to listen to the sounds as many times as they wished.

Results from Experiment I (Table I) indicate that LPLE was able to find a larger number of formants from the vowel data analysed. The maximum number of formants to be modeled by 8th order all-pole modeling is four. This maximum number of formants was indicated in 18 and 49 cases by LP and LPLE, respectively. The smallest number of formants indicated by the two all-pole techniques was two. This minimum amount of formants was indicated by conventional LP in 14 vowels whereas LPLE showed the same amount of formants in only 4 cases. Overall, LPLE showed a larger number of formants than conventional LP in 37 cases, whereas only in 3 sounds were there a larger number of formants indicated by conventional LP than by LPLE.

The results of Experiment II show that listeners pre-

Subject	/a/		/e/		/i/		/o/		/u/		/y/		/ä/		/ö/	
Male 1	2	4	4	4	4	4	3	4	2	4	3	3	4	4	3	4
Male 2	3	4	4	4	4	4	3	4	2	3	3	4	3	3	4	4
Male 3	4	4	4	4	3	3	2	4	2	4	4	4	3	4	3	4
Male 4	4	4	3	4	3	3	3	4	3	4	3	4	3	3	3	3
Male 5	3	4	4	4	4	4	3	4	3	4	2	3	4	4	3	4
Female 1	4	4	3	4	3	3	4	4	3	4	4	4	3	3	3	3
Female 2	3	3	2	4	3	3	3	4	2	3	3	2	3	3	3	3
Female 3	3	3	3	4	2	3	3	2	2	4	3	2	3	4	3	3
Female 4	3	3	3	3	3	3	4	4	3	4	2	3	3	4	3	3
Female 5	3	3	2	2	2	3	4	4	3	4	3	4	2	4	3	3

TABLE I

NUMBER OF FORMANTS INDICATED BY CONVENTIONAL LP AND LPLE ($m = 8$), RESPECTIVELY, FOR EIGHT FINNISH VOWELS PRODUCED BY TEN SUBJECTS.

ferred the quality of the vowels synthesized with LPLE especially in the case when F1 and F2 of the underlying vowel were located at the lower end of the frequency band. The vowel /u/, which is the the vowel characterized by the lowest values of F1 and F2, was perceived by all the listeners to be closer to the original vowel when the synthesis was carried out with LPLE. The vowel /a/ yielded almost the same result, except for one listener who considered the sound synthesized by conventional LP to be of a more natural quality. In the case of the vowel /i/, which has the highest value of the second formant among the vowels, the opinions of the panelists veered away from favoring LPLE: three panelists regarded the perceptual quality of the vowel synthesized by conventional LP to be closer to the original one, three considered the two methods to be of the same quality and two considered the quality given by LPLE to be better.

Two representative examples of the all-pole spectra obtained are shown in Fig. 4. Both figures illustrate how LPLE more clearly indicates the lowest two formants in comparison to conventional LP.

V. CONCLUSIONS

Herein, we have presented a new linear predictive technique which puts more emphasis on the lower frequency range of the speech spectrum. The method is based on first interpreting conventional linear predictors of successive prediction orders with parallel structures using the concept of symmetric linear prediction. The parallel structure of the new method is then obtained by combining a symmetric linear predictor preceded by a low-pass pre-filter from the m'th order prediction to a symmetric predictor preceded by no pre-filter of the $(m - 1)$'th order prediction.

Computation of an m'th order LPLE filter requires a larger number of arithmetic operations than computation
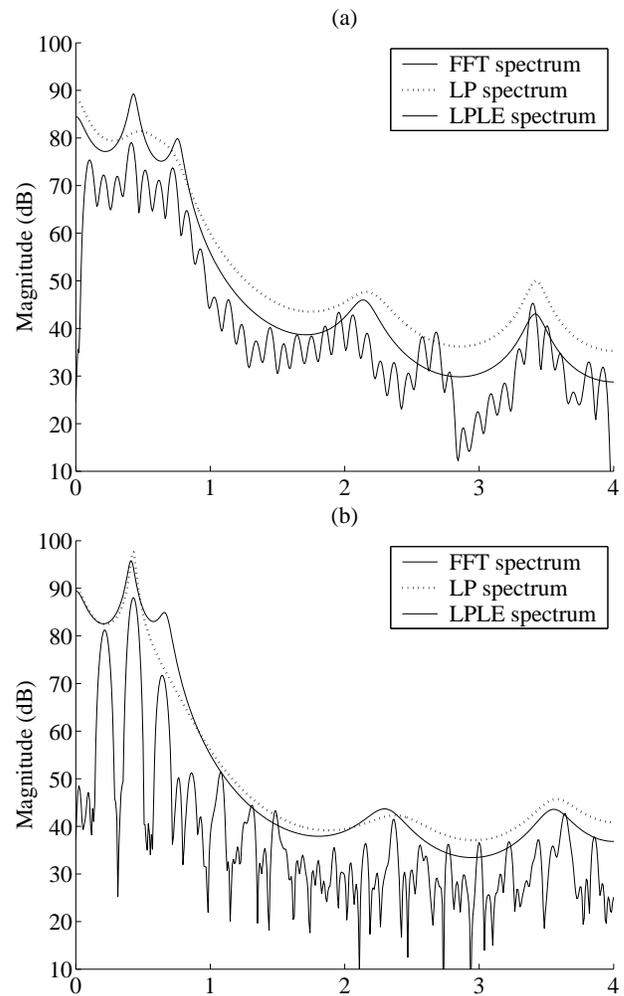
Fig. 4

EXAMPLES OF 8TH ORDER ALL-POLE SPECTRA COMPUTED BY CONVENTIONAL LP (DOTTED LINE) AND BY LPLE (SOLID LINE). (A) MALE SPEAKER, VOWEL /O/, (B) FEMALE SPEAKER, VOWEL /U/. BOTH FFT SPECTRA HAVE BEEN LOWERED BY $10dB$ FOR VISUAL CLARITY.

of the conventional LP filter of the same order. This is especially due to the fact that in determining an LPLE filter of order m with the algorithm described in Section III-A, one needs to inverse two autocorrelation matrices (stages No. 1 and 2 in the algorithm) both of which are of dimensions $(m/2) \times (m/2)$. Moreover, autocorrelation terms needed in stage 1 and 2 are different due to the fact the pre-filter $P_1(z)$ is involved in stage 1 but not in stage 2. However, the number of computations can be reduced by interpreting $B_{1,e}(z)$ in stage 1 and $B_{1,o}(z)$ in stage 2 as the LSP-polynomial $P(z)$ determined for a m'th order and $(m-1)$'th order LP predictor, respectively.

In the current paper, the emphasis has been devoted to presenting the new algorithm, to show that the all-pole filters obtained are stable and to make preliminary experiments with vowel sounds. In principle, it is possible to compare the proposed method with conventional LP of the same prediction order as a part of a real speech coder but that was not in the scope of the present paper. It is, however, worth noticing that there are no principal obstacles to replace the conventional LP with LPLE also inside a speech coding algorithm, because both algorithms define a minimum-phase predictor of the same order with a reasonable computational load. In the preliminary experiments performed in the current study, we focused on the improvement in the modeling of the lowest formants and, therefore, it was a reasonable choice to select vowels as speech data. An extensive evaluation of the new method would also call for using larger speech material and to test, how the performance of LPLE behaves in such speech sounds (e.g., fricatives), which have their energy concentrated on higher frequencies.

The proposed method always yields stable all-pole filters. The method is well-suited to all-pole modeling applications where prediction order needs to be low and where more emphasis needs to be put on the modeling of the lower part of the frequency range. One such area, although not studied in the present survey, is linear predictive analysis of wide-band speech, which calls for using large prediction orders when conventional LP is used. With the method proposed in this paper, however, it is possible to use smaller prediction orders yet obtain all-pole models capable of separating the most important lowest formants.

REFERENCES

[1] −, Recommendation GSM 06.10 "GSM full rate speech transcoding", ETSI, 1992.
[2] −, Digital cellular telecommunications system (Phase 2); Enhanced full rate (EFR) speech transcoding, GSM 06.60 version 4.0.0, ETSI, August 1997.
[3] −, Recommendation G.729-Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), ITU-T, March 1996.
[4] J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE, vol. 63, no. 5, pp. 561–580, April 1975.
[5] J. D. Markel and A. H. Gray, Linear prediction of speech, Springer, Berlin, 2nd edition, 1980.
[6] B. Moore, Ed., Hearing, Academic Press, London, 1995.
[7] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," J. Acoust. Soc. Amer., vol. 87, no. 4, pp. 1738–1751, 1990.
[8] A. Härmä, "Linear predictive coding with modified filter structures," IEEE Trans. Speech Audio Proc., vol. 9, no. 8, pp. 769–777, November 2001.
[9] P. L. Chu and D. G. Messerschmitt, "Frequency weighted linear prediction," in Proc. ICASSP'82, May 1982, pp. 1318 – 1321.
[10] P. L. Chu and D. G. Messerschmitt, "A frequency weighted Itakura-Saito spectral distance measure," IEEE Trans. Acoust., Speech, Signal Processing, vol. 30, pp. 545 – 560, August 1982.
[11] F. K. Soong and B.-H. Juang, "Line spectrum pair (LSP) and speech data compression," in Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'84, San Diego, CA, March 1984, vol. 1, pp. 1.10.1–1.10.4.
[12] P. Stoica and A. Nehorai, "The poles of symmetric linear prediction models lie on the unit circle," IEEE Trans. Acoust. Speech Signal Proc., vol. ASSP-34, pp. 1419–1426, Dec. 1986.
[13] P. Stoica and A. Nehorai, "On linear prediction models constrained to have unit-modulus poles and their use for sinusoidal frequency estimation," IEEE Trans. Acoust. Speech Signal Proc., vol. 36, no. 6, pp. 940–942, June 1988.
[14] H. W. Schüssler, "A stability theorem for discrete systems," IEEE Trans. Acoust. Speech Signal Proc., vol. ASSP-24, no. 1, pp. 87–89, Feb. 1976.
[15] G. A. Mian and G. Riccardi, "A localization property of line spectrum frequencies," IEEE Trans. Speech Audio Proc., vol. 2, no. 4, pp. 536–539, 1994.
[16] H. K. Kim and H. S. Lee, "Interlacing properties of line spectrum pair frequencies," IEEE Trans. Speech Audio Proc., vol. 7, no. 1, pp. 87–91, 1999.
[17] K. K. Paliwal and W. B. Kleijn, "Quantization of LPC parameters," in Speech Coding and Synthesis, W. B. Kleijn and K. K. Paliwal, Eds., pp. 433–466. Elsevier, Amsterdam, 1995.
[18] K. K. Paliwal, "Interpolation properties of linear prediction parametric representations," in Proc. IEEE Speech Coding Workshop, Annapolis, MD, 1995.
[19] K. K. Paliwal, "A study of line spectrum pair frequencies for vowel recognition," Speech Commun., vol. 8, pp. 27–33, 1989.
[20] S. H. Choi, H. K. Kim, and H. S. Lee, "Speech recognition using quantized LSP parameters and their transformations in digital communication," Speech Commun., vol. 30, pp. 223–233, 2000.
[21] C. S. Liu, M. T. Lin, W. J. Wang, and H. C. Wang, "Study of line spectrum pair frequencies for speaker recognition," in Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP'90, 1990, pp. 277–280.
[22] A. V. Oppenheim and R. W. Schafer, Discrete-Time Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1989.
[23] G. Peterson and H. Barney, "Control methods used in a study of the vowels," J. Acoust. Soc. Amer., vol. 24, pp. 175–184, 1952.