ALL-POLE MODELING OF WIDE-BAND SPEECH WITH SYMMETRIC LINEAR PREDICTION

Paavo Alku and Tom Bäckström

Helsinki University of Technology
Lab. of Acoustics and Audio Signal Processing
P.O.Box 3000, Fin-02015 HUT, e-mail: Paavo.Alku@hut.fi

ABSTRACT

A new linear predictive technique, All-pole modeling with Symmetric Linear Prediction (ASLP), is presented. The starting point of the method is an implementation of conventional linear prediction (LP) with a parallel structure, where two *symmetric* linear predictors are combined to pre-filters represented by first order FIRs. Modification of these pre-filters yields the ASLP predictor, which is always minimum phase. Experiments indicate that the new method models the formant structure of wide-band speech more accurately than conventional LP, when the prediction order is smaller than the one required by the sampling frequency.

1. INTRODUCTION

Linear prediction (LP) is a widely used method in speech processing. According to the classical "rule of thumb", the prediction order (denoted by *m* hereafter) of LP analysis should be selected to equal the sampling frequency in kHz added by a small integer [3]. Hence, LP analysis of order m=10 is typically used in compression of telephone band speech sampled with 8 kHz. Processing of wide-band speech, however, calls for using higher prediction orders, which, in turn, increases the amount of side information in LP-based wide-band coders. Therefore, in order to save bits in quantization of the LP parameters of wide-band speech, it would be tempted to use prediction orders somewhat smaller than those given by the "rule of thumb". If, on the other hand, the prediction order is too small, the formant structure of speech becomes poorly modeled. In particular, modeling of the first (F1) and second (F2) formant, which are perceptually among the most important parameters of speech, will deteriorate if LP with a too small prediction order is used for wide-band speech.

This study presents a new linear predictive method, which gives stable all-pole models of wide-band speech with improved modeling of lowest formants, when the prediction order is smaller than the sampling frequency in kHz. The proposed method is described in section 3.1 by first presenting in section 2 three different implementations for conventional LP.

2. BACKGROUND

We will first present shortly three background issues of the new algorithm: conventional LP, the Line Spectral Pairs decomposition, and symmetric linear prediction. For the sake of conciseness, we will assume throughout the study that the prediction order *m* is even. This is because LP models of odd order occur seldom in applications and they can be treated, if needed, similarly.

2.1. Conventional linear prediction

Conventional LP with the prediction order equal to *m* can be presented as follows [2]. Given the signal $x(n)$ and the predictor parameters a_i ($0 \le i \le m$), denoted by $\mathbf{x} = [x(n) \ldots x(n-m)]^T$ and $\mathbf{a} = [a_0 \ldots a_m]^T$, 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 $E[x^2(n)] = \mathbf{a}^T\mathbf{R}\mathbf{a}$ subject to the constraint a_0=1 or equivalently $\mathbf{a}^T\mathbf{b}$=1, where $\mathbf{b}$= $[1\ 0 \ldots 0]^T$. The solution can be written as: $\mathbf{R}\mathbf{a}=\sigma^2\mathbf{b}$, where σ^2 denotes the energy of the residual. In the following, we will denote the transfer function of the predictor given by conventional LP by $A(z)$.

2.2. LSP (Line Spectral Pairs) decomposition

Given an LP predictor $A(z)$, the Line Spectral Pair (LSP) polynomials are defined as [5]: $P(z) = A(z) + z^{-m-1}A(z^{-1})$, $Q(z) = A(z) - z^{-m-1}A(z^{-1})$. Predictor $A(z)$ can be obtained from the LSP polynomials simply as: $A(z) = 1/2[P(z) + Q(z)]$.

The LSP decomposition is widely used in quantization of LP parameters [4]. It is known that zeros of $P(z)$ and $Q(z)$ are always on the unit circle and, for even values of *m*, $P(z)$ and $Q(z)$ have trivial roots located at z=-1 and z=+1, respectively. Furthermore, if the roots of P(z) and Q(z) interlace, the all-pole filter $1/A(z)$ is stable.

2.3. Symmetric linear prediction

Symmetric linear prediction [6,7] is based on the predictor polynomial defined as:

$$B(z) = 1 + b_1 z^{-1} + \ldots + b_{m/2-1} z^{-m/2+1} + b_{m/2} z^{-m/2} + b_{m/2-1} z^{-m/2-1} + ..b_1 z^{-m+1} + z^{-m} \quad (1)$$

In this structure, a predictor of order *m* can be defined from *m/2* coefficients b_i ($1 \le i \le m/2$) because of the symmetry of the impulse response. The optimal predictor is obtained from the following normal equations:

$$\sum_{k=1}^{m/2-1} b_k\Big(R(k-j)+R(k-m+j)\Big)+b_{m/2}R(j-m/2) \qquad (2)$$
$$=-R(j)-R(j-m),\; j\in[1,m/2]$$

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

It has been proved that roots of B(z) are always on the unit circle [6]. Interestingly, there is a close connection (although not well known) between symmetric LP and the LSP decomposition: the LSP polynomials (excepting the points z=+/- 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., residual is computed using symmetric linear prediction) [7].

By summarizing the background issues, we can now express conventional LP using three different implementations as shown in Fig. 1.

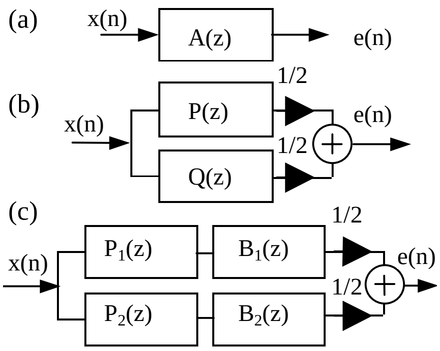

Fig.1 Different implementations of conventional LP, x(n) and e(n) denote the input and the residual, respectively.

(a) Conventional implementation with predictor A(z).
(b) Implementation based on the LSP decomposition.
(c) Implementation based on symmetric LP. Transfer functions of symmetric linear predictors $B_1(z)$ and $B_2(z)$ obey the structure given in Eq. 1 and they are optimized (using Eq. 2) from x(n) filtered through pre-filters $P_1(z)$ and $P_2(z)$, respectively. Transfer functions of pre-filters are: $P_1(z)=1+z^{-1}$ and $P_2(z)=1-z^{-1}$.

Normally, conventional linear prediction is computed by inverse filtering x(n) using the optimal predictor A(z) (Fig. 1(a)). With the LSP decomposition, however, it is possible to implement the same filtering using the parallel structure shown in Fig. 1(b). This implementation requires more computations than the one shown in Fig. 1(a). However, it serves as an intermediate step to the third implementation of conventional LP shown in Fig. 1(c). In this alternative, the trivial roots of the LSP polynomials have been separated and, consequently, both P(z) and Q(z) are expressed as cascades, where two *symmetric* LP predictors (with transfer functions denoted by $B_1(z)$ and $B_2(z)$) are defined from x(n) filtered through *pre-filters* represented by simple FIRs ($1+z^{-1}$ and $1-z^{-1}$ for P(z) and Q(z), respectively). Hence, the transfer function of the conventional LP predictor computed by the structure shown in Fig. 1(c) can be expressed as

$$A(z) = 1/2\,(\,B_1(z)P_1(z) + B_2(z)\,P_2(z)\,) \qquad (3)$$

where $B_1(z)$ and $B_2(z)$ are transfer functions of the symmetric linear predictors computed from x(n) filtered through pre-filters $P_1(z) = 1+z^{-1}$ and $P_2(z)=1-z^{-1}$.

Implementation of conventional LP using the structure shown in Fig. 1(c) is, again, inferior to the one shown in Fig.1(a) in terms of the computational load. However, this interpretation of conventional LP makes possible a straightforward modification, which introduces the new algorithm described next in section 3.1.

3. METHOD

3.1. Algorithm

When interpreting conventional LP using Eq. 3 it is worth noticing that the zeros of $P_1(z)$ and $P_2(z)$ are at z=-1 and z=+1, respectively, which implies that the amplitude responses of these two pre-filters at the low and high end of the frequency range are considerably different. Therefore, the symmetric linear predictor $B_1(z)$ in the upper branch of Fig.1(c) will most likely not locate any of its roots at the highest frequencies 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 $B_2(z)$ 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 large distance between a root of $B_1(z)$ (on the unit circle) and its counterpart of $B_2(z)$ (also on the unit circle), the corresponding root of the final LP predictor (inside the unit circle) will be at a distance from the unit circle. (This follows from properties between LSP roots and LP.). Hence, the spectral model given by LP will most likely not show a strong resonance in the frequency range in question.

The proposed method, All-pole modeling with Symmetric Linear Prediction (ASLP), is based on a straightforward modification of the structure shown in Fig. 1(c) in order to obtain all-pole models, which emphasize more the lowest and highest frequencies of the input signal. The flow graph of ASLP is shown in Fig. 2. In this method, pre-filters $P_1(z)$ and $P_2(z)$ of Fig. 1(c) are replaced by first-order pre-filters, the transfer functions of which are $P_{ASLP,1}(z) = 1+az^{-1}$ and $P_{ASLP,2}(z) = 1-az^{1}$ in the lower and upper branch, respectively. If $|a|<1$, the amplitude response of the pre-filter $P_{ASLP,1}(z)$ will introduce smaller attenuation on the input spectral components at the highest spectral range than $P_1(z)$. Similarly, $P_{ASLP,2}(z)$ will not attenuate the lowest frequency range as much as $P_2(z)$. Therefore, we would expect the roots of the symmetric linear predictors to become closer to each other at the two ends of the frequency range. Consequently, the final all-pole model should indicate more prominent resonances especially at the lower frequencies of wide-band vowels. Selection of the parameter *a* in the current study is described in section 4.

In order to guarantee the minimum phase property of the final predictor, the two symmetric predictors $B_{ASLP,1}(z)$ and $B_{ASLP,2}(z)$ have to be cascaded with pre-filters $P_1(z)$ and $P_2(z)$, respectively. (This can be proved using properties of symmetric polynomials as shown in [1], but it is beyond the scope of this article.) Hence, the transfer function of the predictor given by ASLP can be expressed as:

$$A_{ASLP}(z) = 1/2\ (B_{ASLP,1}(z)P_1(z) + B_{ASLP,2}(z)\ P_2(z)) \quad (4)$$

where $B_{ASLP,1}(z)$ and $B_{ASLP,2}(z)$ are symmetric linear predictors defined from x(n) filtered through pre-filters $P_{ASLP,1}(z)$ and $P_{ASLP,2}(z)$, respectively. $P_1(z)$ and $P_2(z)$ are transfer functions of the pre-filters used in the implementation of the conventional LP with $P_1(z) = 1+z^{-1}$ and $P_2(z)=1-z^{-1}$.

It is worth noticing that the order of the predictor determined with this new approach will be equal to *m*, because the highest terms in products $B_{ASLP,1}(z)P_1(z)$ and $B_{ASLP,2}(z)P_2(z)$ cancel each other when summed together in Eq. 4. From the point of view of speech coding, the proposed algorithm is similar to conventional LP, because it makes possible presenting a predictor of order *m* with 2*m* roots located on the unit circle. (In conventional LP, these are the roots of P(z) and Q(z), ie., the LSPs, while in ASLP these are the roots of the two symmetric LP polynomials $B_{ASLP,1}(z)$ and $B_{ASLP,2}(z)$). Hence, same amount of bits is needed in quantization of the *m*th order predictor in conventional LP and in ASLP.

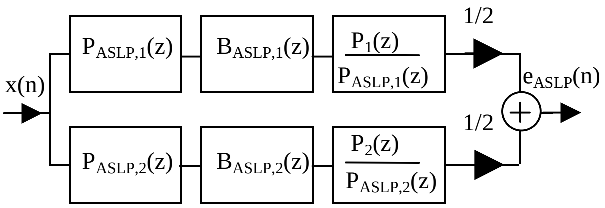

Fig.2 Flow graph of the predictor in ASLP, x(n) and $e_{ASLP}(n)$ denote the input and residual, respectively. Symmetric linear predictors $B_{ASLP,1}(z)$ and $B_{ASLP,2}(z)$ are defined (using Eq. 2) from signal x(n) filtered through pre-filters $P_{ASLP,1}(z)$ and $P_{ASLP,2}(z)$, respectively. The ASLP predictor is obtained (Eq. 4) by cascading the symmetric linear predictors with pre-filters $P_1(z)$ and $P_2(z)$. The transfer functions of the pre-filters are: $P_1(z)=1+z^{-1}$, $P_2(z)=1-z^{-1}$, $P_{ASLP,1}(z)= 1+az^{-1}$, $P_{ASLP,2}(z)=1-az^{-1}$.

4. EXPERIMENTS

As speech material we recorded vowels (/a/ and /o/) produced by five female and five males subjects. The data were digitized using a sampling frequency of 22.050 kHz. In the computer, the sounds were high-pass filtered in order to remove low-frequency fluctuations picked up in the recordings. The final analysis bandwidth of the vowels was between 250 Hz and 11.025 kHz.

Both LP and ASLP were computed with the prediction order *m*=16 and the frame size of 1000 samples (= approx. 45 ms). The autocorrelation terms were computed using Hamming windowing. The all-pole models computed by ASLP might sometimes show resonances that are too narrow (This happens typically in modeling of the first formant.) To alleviate this problem, the impulse response of the predictor computed by ASLP was windowed with the exponential window $w(n) = 0.99^n$, $0 \le n \le m$.

Comparison of the all-pole spectra were computed using the Spectral Distortion (SD) [4]. This distortion measure is defined as follows:

$$SD = \sqrt{\frac{1}{f_u - f_l}\int_{f_l}^{f_u}[10\log_{10} P_1(f) - 10\log_{10} P_2(f)]^2\, df} \quad (5)$$

where $P_1(f)$ and $P_2(f)$ denote the all-pole spectra to be compared and f_l and f_u denote the lower and upper limit, respectively, of the frequency range to be analyzed. (In computation of SD, the gains of the all-pole filters were adjusted so that their energies equaled the energy of the analyzed speech sound.) Using SD, the parameter *a* of pre-filters $P_{ASLP,1}(z)$ and $P_{ASLP,2}(z)$ in ASLP was selected as follows. Firstly, ASLP was computed for each vowel with 20 different values of *a*, by varying the value of *a* from 0.05 to 1.0 (i.e., conventional LP) with the step of 0.05. Secondly, higher order LP analyses (*m*=26) were computed for all the vowels. Each 26th order all-pole spectrum was then compared to the twenty all-pole spectra given by ASLP using SD. In order to assess the modeling of the lowest two formants, SD was computed with f_l=250 Hz and f_u=1500 Hz. It was found that the value of SD was smallest for most of the sounds analyzed when *a* equaled 0.45. Hence, we set *a*=0.45 and this value was kept constant in all the further experiments.

5. RESULTS

ASLP was compared to conventional LP using two approaches. Firstly, conventional LP analysis of a high prediction order (*m*=26) was computed for the vowels /a/ and /o/ produced by a male speaker. The obtained all-pole filters were then excited by an impulse train in order to create synthetic vowels with known formant structures. Five versions of both vowels were produced by varying the fundamental frequency (F0) from 100 Hz to 300 Hz with the step of 50 Hz. Finally, each synthetic vowel was analyzed by conventional LP and ASLP of order *m*=16 and the obtained all-pole spectra were compared to those of the 26th order all-pole spectra used in the synthesis of the vowels. This comparison was again done by using SD as described in section 4. The data obtained are shown in Table 1. Secondly, vowels of all speakers were analyzed by using conventional LP and ASLP of order *m*=16. The performance of LP and ASLP was then compared by extracting the first and the second formant from each all-pole spectrum. The formant was identified as a local maximum in the all-pole spectrum in the frequency range, where F1 and F2 of the vowels /a/ and /o/ are known to locate.

The SD values computed from the synthetic vowels show that the 16th order all-pole spectra given by ASLP were closer to the 26th order LP spectra than those given by the 16th order conventional LP analysis. As shown by the data of Table 1, ASLP yielded smaller SD values for all the ten vowels analyzed. These data reflect especially the more accurate modeling of the lowest two formants by the ASLP in comparison the conventional LP analysis of the same prediction order.

Improved modeling of the lowest two formants by ASLP was also supported by the number of the formants found. Both of the two all-pole modeling techniques indicated the first formant in all 20 sounds analyzed, but conventional LP analysis failed to indicate the second formant in eight cases, whereas ASLP found both of the lowest two formants for all the sounds analyzed. Fig. 3 shows two examples demonstrating the improved modeling of the lowest two formants by ASLP.

	Vowel /a/		*Vowel /o/*	
F0	SD_{LP}	SD_{ASLP}	SD_{LP}	SD_{ASLP}
100	1.15	0.55	0.88	0.54
150	1.20	0.50	1.50	0.95
200	1.26	0.87	1.26	1.18
250	1.16	0.92	1.69	1.07
300	1.18	0.63	1.53	1.05

Table 1. Spectral distortion (SD, in dB) computed from the synthetic vowels with different values of the fundamental frequency (F0, in Hz). SD was computed using Eq. 5 with the following parameters: $P_1(f)$ was the 26th order all-pole spectrum used in the synthesis of the sounds, $P_2(f)$ was the 16th order all-pole spectrum given either by conventional LP (SD_{LP}) or by ASLP (SD_{ASLP}). The analysis was computed over the frequency range, where the lowest two formants are located (f_l=250 Hz and f_u=1500 Hz).

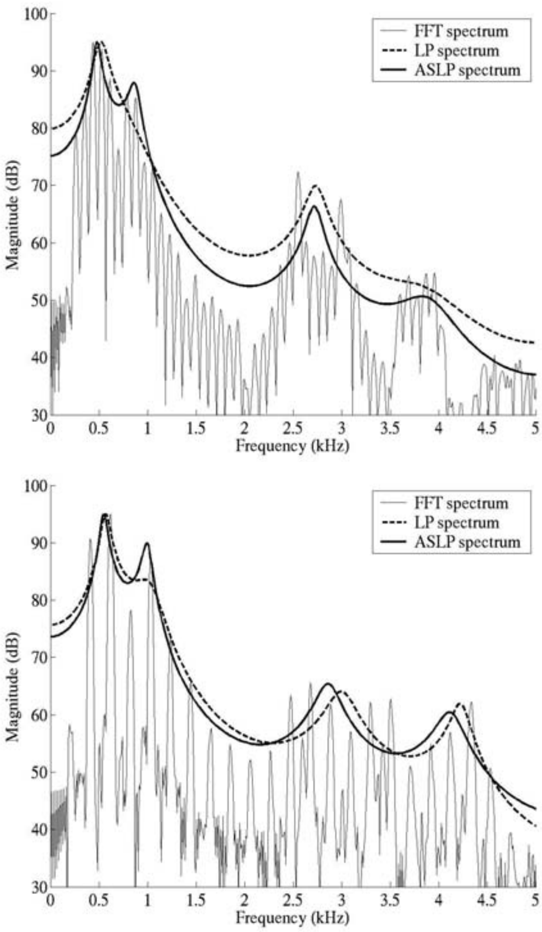

Fig. 3 Examples of 16th order all-pole spectra computed for the vowel /a/ by conventional LP (dotted line) and by ASLP (solid line). Analysis bandwidth was up to 11.025 kHz, but, in order to make figures clear, spectra are shown only up to 5 kHz. Upper graph: male speaker, lower graph: female speaker.

6. SUMMARY

We have presented a new linear predictive technique that gives stable all-pole filters to model wide-band speech spectra. The method is based on an implementation of a linear predictor as a parallel structure, where two symmetric linear predictors are determined by filtering the input through two pre-filters. By applying pre-filters different from those used in the corresponding implementation of conventional LP, the proposed method makes possible emphasizing especially the lowest frequencies of wide-band speech prior to the computation of the symmetric linear predictors. Consequently, the poles of the all-pole filter obtained are more likely to locate at low frequencies than in conventional LP. This phenomenon was corroborated by the experiments which showed that the proposed ASLP method models the first and second formant of wide-band vowels more accurately then conventional LP of the same prediction order.

7. REFERENCES

[1] Bäckström, T., Alku, P. "On the stability of constrained linear predictive models," submitted to ICASSP'03.

[2] Makhoul, J. "Linear prediction: A tutorial review," Proc. IEEE, Vol. 63, No. 4, pp. 561-580, April 1975.

[3] Markel, J.D., Gray, A.H., Jr. *Linear Prediction of Speech*, pp. 154-156., Springer New York., 1976.

[4] Paliwal, K.K., Kleijn, W.B. "Quantization of LPC parameters," in Speech Coding and Synthesis, W.B. Kleijn and K.K. Paliwal, Eds., Elsevier, Amsterdam, 1995.

[5] Soong, F.K, Juang, B-H. "Line spectrum pair (LSP) and speech data compression," in Proc. ICASSP'84, San Diego, CA, pp.1.10.1-1.10.4.

[6] Stoica, P., Nehorai, A. "The poles of symmetric linear prediction models lie on the unit circle," IEEE Trans. on Acoust.Speech and Signal Proc., Vol. 34, pp. 1344-1346, Oct.1986.

[7] Stoica, P., Nehorai, A. "On linear prediction models constrained to have unit-modulus poles and their use for sinusoidal frequency estimation," IEEE Trans. on Acoust.Speech and Signal Proc., Vol. 36, No. 6, pp. 940-942, June 1988.