ALL-POLE MODELING OF WIDE-BAND SPEECH USING WEIGHTED SUM OF THE LSP POLYNOMIALS

Paavo Alku and Tom Bäckström

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

ABSTRACT
The autocorrelation function of the all-pole filter given by the conventional linear prediction (LP) matches exactly the autocorrelation function of the input signal between indices 0 and m, when the prediction order equals m. This study describes a recently developed technique, Weighted-Sum Line Spectrum Pair (WLSP), where an all-pole filter is defined by using a sum of weighted LSP (Line Spectrum Pair) polynomials. WLSP yields a stable all-pole filter of order m, whose autocorrelation function coincides to that of the input between indices 0 and m − 1. By sacrificing the exact matching of the autocorrelations at index m, WLSP models the autocorrelation of the input at the indices above m more accurately than conventional LP. In the current paper, the performance of WLSP in spectral modelling of wide-band speech is analysed. It is shown that WLSP yields all-pole spectra that model the formant structure of vowels more accurately than conventional LP of the same prediction order.

1. INTRODUCTION
Linear prediction (LP) is one of the most widely used methods of speech processing. Especially in low bit rate speech coding the role of LP is most essential and majority of the speech coders standardised during the past two decades are based on linear prediction [1, 2]. Among the different variations of LP, the autocorrelation method of linear prediction is the most popular. In this method, a predictor (an FIR of order m) is determined by minimising the square of the prediction error, the residual, over an time interval which is, in principle, infinite [3]. Popularity of the conventional autocorrelation method of LP is explained by its ability to compute with a reasonable computational load a stable all-pole model for the speech spectrum, which is accurate enough for most applications when presented by a few parameters. The performance of LP in modelling of the speech spectrum can be explained by the autocorrelation function of the all-pole filter, which matches exactly the autocorrelation of the input signal between indexes 0 and m when the prediction order equals m [4]. It should be noticed, however, that in the computation of the conventional LP analysis of order m no information is available about the autocorrelation function of the input beyond time index m.

The Line Spectrum Pair (LSP) decomposition is among the methods developed for presenting the LP information (i.e., data required to express the transfer function of the m’th order LP predictor) [5]. In this technique, the minimum phase predictor polynomial computed by the autocorrelation method of linear prediction is split into a symmetric and an antisymmetric polynomial. It has been proved that the roots of these two polynomials, the LSPs, are located interlaced on the unit circle, if the original LP predictor is minimum phase [6]. Furthermore, it has been shown that LSPs behave well when interpolated [7]. Due to these properties, the LSP decomposition has become the major technique in quantisation of LP information and it is used in various speech coding algorithms [1, 2].

In [8], we presented a new linear predictive algorithm, Weighted-sum Line Spectrum Pair (WLSP), which yields an all-pole filter of order m to model the speech spectrum. In contrast to the conventional linear prediction, the proposed
algorithm (with its prediction order equal to \( m \)) takes advantage of the autocorrelation of the input signal also beyond time index \( m \) in order to obtain a more accurate all-pole model for the speech spectrum. WLSP utilises the LSP decomposition in a manner different from that typically used in speech coding: the LSP decomposition is not computed in order to quantise the LP information but rather as a computational tool, using which stable all-pole filters with the proposed autocorrelation matching property are defined.

The current study focuses on all-pole modelling of wide band speech. More specifically, we were interested in comparing the performance of the new WLSP method to that of the conventional LP-analysis in modelling of the formant structure of wide-band speech (bandwidth 8 kHz, sampling frequency 16 kHz). In both analyses, we used the order of prediction \( m = 16 \) which is applied in the AMR Wide-band speech codec of the forthcoming 3rd generation mobile communication system [1]. Using the conventional LP-analysis of order \( m = 16 \) might result, unfortunately, in poor modelling of those formants of wide-band speech that are located close to each other at low frequencies (see Fig. 1). Therefore, we aimed at analysing whether the new WLSP method could distinguish these perceptually most important formants more accurately.

2. METHODS

2.1. Linear prediction and LSP

The conventional LP predictor of order \( m \) for a signal \( x(n) \) as given by [3] is

\[
A_m(z) = 1 + \sum_{i=1}^{m} a_i z^{-i}.
\]

(1)

The coefficient vector \( a = (a_1, \ldots, a_m, 0) \), where \( a_0 = 1 \), can be solved from the normal equations \( Ra = \sigma^2 [1, 0, \ldots, 0]^T \).

The autocorrelation matrix \( R \) is defined as the expected value of correlation \( xx^T \), that is, \( R = E[xx^T] \), where signal \( x \) is assumed wide-sense stationary and the residual energy is \( \sigma^2 = a^T Ra \).

The symmetric and antisymmetric LSP polynomials are defined as \( P(z) = A_m(z) + z^{-m-1}A_m(z^{-1}) \) and \( Q(z) = A_m(z) - z^{-m-1}A_m(z^{-1}) \), respectively [6]. By defining a zero-extended vector \( \hat{a} = [a^T, 0]^T \) (where \( a \) is the LP coefficient vector), the LSP polynomials can be defined equivalently in matrix form as

\[
p = \hat{a} + \hat{a}_R \quad \text{and} \quad q = \hat{a} - \hat{a}_R,
\]

where subscript \( R \) denotes reversal of rows.

2.2. Weighted sum of the LSP polynomials

In [8], we proposed the WLSP method based on the predictor polynomial \( D(z, \lambda) \) defined as

\[
D(z, \lambda) = \lambda P(z) + (1 - \lambda)Q(z),
\]

(3)

that is, \( D(z, \lambda) \) is a weighted sum, with weighting parameter \( \lambda \), of the LSP polynomials \( P(z) \) and \( Q(z) \) in the polynomial space. It is worth noticing that in the open interval \( \lambda \in (0, 1) \) polynomial \( D(z, \lambda) \) is minimum-phase [8]. Further, setting \( \lambda = \frac{1}{2} \) yields the original LP polynomial \( A_m(z) \). Choosing the value of \( \lambda \in (0, 1) \) which minimises the residual energy (the Levinson-Durbin solution [9]) yields the LP polynomial \( A_{m+1}(z) \). In the interval \( \lambda \in (0, 1) \) the autocorrelation function of the inverse of the predictor \( D^{-1}(z, \lambda) \) fits exactly the first \( m - 1 \) values of the autocorrelation of the input [8].

2.3. Algorithm

The weighted sum of the LSP polynomials (Eq. 3) together with the autocorrelation properties of \( D^{-1}(z, \lambda) \) serves as the basis for the proposed WLSP filters. Given an input signal \( x(n) \), defining a stable WLSP all-pole filter of order \( m \) comprises the following stages. (The time index of the autocorrelation function is denoted by \( i \).

1. Calculate LP polynomial \( A_{m-1}(z) \) for signal \( x(n) \) using conventional LP with the autocorrelation criterion. (Notice that LP analysis of order \( m \) yields LSP polynomials of order \( m + 1 \) [6]. Therefore, in order to compute an \( m \)th order WLSP filter, the orders of \( P(z) \) and \( Q(z) \) in Eq. 3 have to be equal to \( m \), which implies that the order of the LP analysis is \( m - 1 \).)

2. Construct the LSP polynomials \( P(z) \) and \( Q(z) \) from \( A_{m-1}(z) \) as defined in [6].
3. Identify the first peak of the autocorrelation function $R$ of the input signal beyond $m$. Define $I$ as the position of this peak.

4. Using Eq. 3, define an all-pole filter $D^{-1}(z, \lambda)$ of order $m$ with an additional parameter $\lambda$. Optimise the all-pole filter by searching for the value of $\lambda$ that minimises the absolute difference between the (normalised) autocorrelations of $x(n)$ and $D^{-1}(z, \lambda)$ for $m \leq i \leq I$. (Notice that autocorrelations of $x(n)$ and $D(z, \lambda)$ are equal for $0 \leq i < m$, when the order of $D(z, \lambda)$ equals $m$.) See Fig. 2 for illustration.

Note that the procedure above is slightly different from the one presented in [8], because we now use an adaptive length of the optimisation window.

3. RESULTS

WLSP was compared to conventional LP by analysing two Finnish vowels (/a/ and /o/) produced by two female and four male speakers. The two linear predictive analyses were computed by using a prediction order $m = 16$, a 20 ms Hamming window and a sampling frequency of 16 kHz. Comparison of WLSP and LP in modelling of formants was done by using the following procedure. Given an all-pole spectrum (in dB), the formant peak was defined as the local maximum of the spectrum. The spectral valley was then defined as the local minimum of the spectrum following this peak. The level difference of these two spectral components, denoted by $L_{\text{diff}}$, was computed for the lowest two formants to characterise the dynamics of the all-pole spectrum in the vicinity of these resonances. $L_{\text{diff}}$ given by LP was then subtracted from $L_{\text{diff}}$ yielded by WLSP. This difference $\Delta L = L_{\text{diff,WLSP}} - L_{\text{diff,LP}}$ is positive in the case WLSP yields a deeper spectral valley between two formants than LP. The value of $\Delta L$ is shown in Figure 3 for all the vowels analysed. It can be noticed from this figure that WLSP yields in all the cases (except for the vowel /o/ produced by female 1) more distinguished spectral valleys between the lowest formants. Value of $\Delta L$ averaged over all the utterances analysed was 1.76 dB and 1.84 dB for the first and second formant, respectively. Moreover, it was found that WLSP modelled in 7 cases out of the total of 12 cases a larger number of formants than LP and just once one formant less. Figures 4 to 6 show examples of all-pole spectra obtained.
Fig. 6. Model performance \((m = 16)\) in the frequency domain for female vowel /o/.

4. CONCLUSIONS

In this article, we have studied a recently developed all-pole modelling technique, Weighted-Sum Line Spectrum Pair (WLSP) in processing of wide-band speech. With this technique, a stable all-pole filter of order \(m\) is defined using the weighted sum of the LSP polynomials. The autocorrelation of the all-pole model matches exactly the \(m - 1\) first autocorrelation values of the input signal, but the \(m'\)th value is sacrificed in order to get improved matching in the upper autocorrelation range.

Preliminary experiments show that WLSP yields, in comparison to conventional LP of the same prediction order, spectral models with larger level differences between formant peaks and spectral valleys.

5. REFERENCES


