All-Pole Modelling of Mixed Excitation Signals

Peter Kabal[1] *and Bastiaan Kleijn*[2]

[1] Dept. Electrical & Computer Engineering
McGill University
Montreal, Canada

[2] Dept. Speech, Music and Hearing
KTH (Royal Institute of Technology)
Stockholm, Sweden

Abstract

Conventional Linear Prediction (LP) techniques can fail to adequately model speech spectra when the model order is too low and/or when the input is periodic (voiced speech). In this paper, we view the LP modelling problem as a correlation matching problem. We introduce a correlation matching criterion which models the signal as a filtered mixture of a noise-like excitation and a periodic excitation. As such it is an extension of the Discrete All-Pole (DAP) modelling approach. The new technique provides a means to generate LP spectra that evolve more smoothly from frame to frame even when the excitation signal has a periodic component with changing period.

1 Introduction

Conventional LP analysis often produces spuriously varying spectra when the LP analysis order is less than the order of the process which generated the data samples and/or when the pitch harmonics interact with the spectral peaks (formants). Conventional LP procedures match the data correlations, but only for the first few correlation lags. By extending the analysis to include more correlation lags, we can get compromise matches which lead to better overall power spectral fits, particularly if the LP analysis order is too small to adequately model the data spectra.

For periodic inputs, the Discrete All-Pole (DAP) model introduced by El-Jaroudi and Makhoul [1] provides a framework for modelling discrete spectra. DAP analysis has been applied to spectral modelling in Multiband Excitation (MBE) [2] and Sinusoidal Transform Coding (STC) [3] coding of speech. In this paper, we extend the DAP framework to handle mixed excitation signals and to incorporate more correlation values.

A model of speech production consisting of an excitation signal driving an all-pole model is shown in Fig. 1. The conventional LP formulation attempts to minimize the prediction error. Another viewpoint is that the task of LP analysis is to identify the all-pole filter used to synthesize the speech.

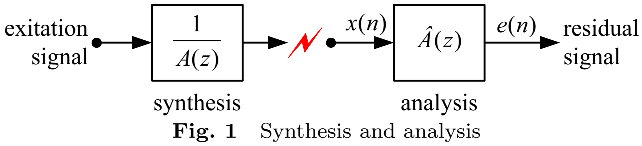

Fig. 1 Synthesis and analysis

Consider a white noise input excitation signal. The filter output data is observed over a long enough interval to obtain good correlation estimates. If the LP analysis order matches that of the synthesis filter, choosing the coefficients of $\hat{A}(z)$ to minimize the energy of the prediction residual leads to the condition that $\hat{A}(z) = A(z)$, i.e., the analysis has identified the synthesis filter parameters. However, for other conditions such as periodic excitation, short observation intervals, and additive noise, ordinary LP analysis may give poor matches.

1.1 Correlation matching properties

The conventional LP (autocorrelation) solution can be expressed using the augmented normal equations ($N_p + 1 \times N_p + 1$),

$$\boldsymbol{R}_{\mathbf{x}} \boldsymbol{a} = \sigma_e^2 \boldsymbol{e}_{\mathbf{0}}, \tag{1}$$

where $\boldsymbol{a}^T = [1, -\mathbf{c}^T]$ and $\boldsymbol{e}_{\mathbf{0}}$ is the unit vector, $\boldsymbol{e}_{\mathbf{0}}^T = [1, 0, \ldots, 0]$. The Toeplitz matrix $\mathbf{R}_{\mathbf{x}}$ is completely determined by the correlation values $r_{xx}(0), \ldots, r_{xx}(N_p)$. The energy of the input signal, $r_{xx}(0)$, scales the output of the prediction error filter. The set of linear equations provides a one-to-one mapping between the set of $N_p + 1$ correlation values and the N_p predictor coefficients together with the prediction residual energy σ_e^2.

The scaled all-pole synthesis filter is,

$$H(z) = G/\hat{A}(z). \tag{2}$$

The correlation of this filter satisfies,

$$r_{hh}(n) = \sum_{k=1}^{N_p} c_k \, r_{hh}(n-k) + G h(-n). \tag{3}$$

or in vector-matrix notation,

$$\boldsymbol{R}_{\mathbf{h}} \boldsymbol{a} = G^2 \boldsymbol{e}_{\mathbf{0}}, \tag{4}$$

where $\boldsymbol{R}_{\mathbf{h}}$ is an $N_v + 1$ by $N_p + 1$ Toeplitz matrix. For $N_v = N_p$, the filter correlations satisfy the same equations as the data correlations. This is the correlation matching property,

$$r_{hh}(k) = r_{xx}(k), \qquad 0 \le k \le N_p. \tag{5}$$

1.2 Extended correlation match

In practice, finite length windows are used to estimate the correlation values. As an example, a white Gaussian excitation is applied to a fixed 12'th order all-pole synthesis filter.[1] The analysis is 8'th order using a 240 sample Hamming window. Figure 2 shows a sequence of estimates.

[1]The first 13 correlation values are 24.3764, 17.3379, 2.4673, −8.1612, −8.0057, −0.7891, 4.2354, 1.8711, −5.0436, −10.5401, −11.4046, −8.6134, and −4.7546. The spectrum appears in Fig. 4.

There are two effects here. First the 8'th order filter cannot completely model the input. Second, the formant peaks change due to the variations in the correlation estimates.

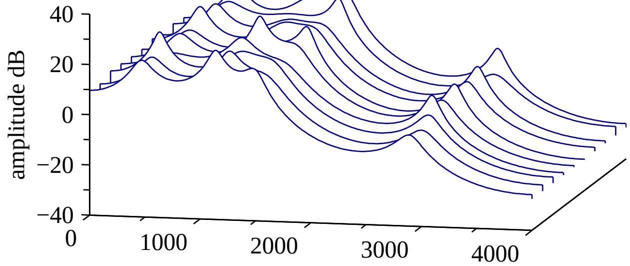

Fig. 2 Conventional LP match (8'th order) spectral fits for noisy correlation estimates (240 sample Hamming window).

Choosing a longer 320 sample window does not improve the situation much. Using multi-taper analysis [4] with approximately the same spectral resolution ($NW = 2$) as the 240 sample Hamming window also does not help much in reducing the variability.

Consider trying to match more correlation values [5, 6],

$$\varepsilon_c^2 = \min_{\boldsymbol{a},G} \sum_{k=0}^{N_v-1} w_k \left(r_{xx}(k) - r_{hh}(k)\right)^2. \tag{6}$$

Minimizing this error criterion leads to a non-linear set of equations. Experiments show that convergence is troublesome and that many local minima exist. In some cases, local minima with nearly equal errors give very different spectra.

As an alternative, we can extend the number of equations in Eq. (1),

$$\varepsilon_e^2 = \min_{\boldsymbol{a},G} \|\boldsymbol{W}^{1/2}(\boldsymbol{R}_{\mathbf{x}}\boldsymbol{a} - G^2\boldsymbol{e}_{\mathbf{0}})\|^2, \tag{7}$$

where $\boldsymbol{W}$ is a diagonal matrix of weights. The matrix $\boldsymbol{R}_{\mathbf{x}}$ is now $N_v + 1$ by $N_p + 1$. This formulation is known as the overdetermined normal equations method [6] or the model equation error method [7]. Minimizing the mean-square equation error leads to a set of linear equations. Since Eq. (7) uses higher order correlations, the match will differ from that of a conventional LP solution.

Figure 3 shows the match for the same set of data considered earlier. Four extra equation terms were used.[2] Note that the formant peaks are now more consistently represented. The emphasis has shifted to better match the higher amplitude formants.

2 Periodic Signals

If the observed signal $x(n)$ is periodic, the correlation values calculated from the data are also periodic,

$$r_{xx}(n) = r_{xx}(n - lP), \tag{8}$$

where P is the integer-valued period. This aliased correlation is what would be calculated from the observed data.

Applying conventional LP analysis to these aliased correlation values gives a different set of predictor coefficients

[2] The weight values used are 1, 0.9998, 0.9969, 0.9847, 0.9533, 0.8931, 0.8011, 0.6850, 0.5603, 0.4431, 0.3430, 0.2628, 0.2011.

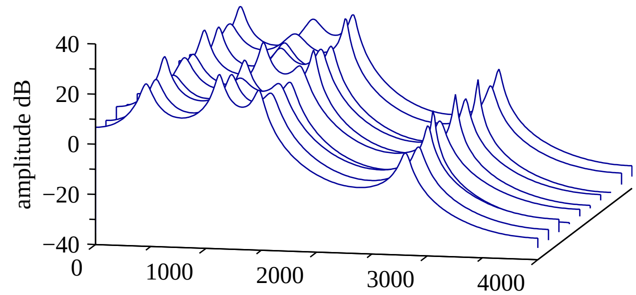

Fig. 3 Overdetermined equation (8'th order) spectral fits for noisy correlation estimates (240 sample Hamming window).

than would have been calculated if the unaliased correlation values were available. The predictor coefficients do not match the coefficients of the all-pole filter that generated the observed periodic signal.

2.1 Discrete all-pole modelling

El-Jaroudi and Makhoul [1] suggested an iterative procedure which they term Discrete All-Pole (DAP) modelling to handle general discrete spectra. Here we focus on discrete spectra generated by periodic sequences. Recall the one-to-one mapping between the predictor coefficients and the correlation values $r_{hh}(n)$. We can form a time-aliased version of the filter response correlations (period P) $\tilde{r}_{hh}(n)$. The error between the (periodic) values $r_{xx}(n)$ and $\tilde{r}_{hh}(n)$ can be used to iterate the filter coefficient values. We develop an approach which can be used to match the time-aliased correlations without explicit calculation of the correlation values $\tilde{r}_{hh}(n)$.

The aliased impulse response of the filter is the response to a periodic train of unit pulses,

$$\tilde{h}(n) = \sum_{l=-\infty}^{\infty} h(n - lP). \tag{9}$$

Note that $\tilde{h}(n)$ is the response to a non-causal input sequence and hence is non-zero for negative n. The correlation function for $\tilde{h}(n)$ is periodic

$$\tilde{r}_{hh}(k) = \sum_{l=-\infty}^{\infty} r_{hh}(k - lP). \tag{10}$$

For a periodic impulse train input, the relationship corresponding to Eq. (3) is

$$\tilde{r}_{hh}(n) = \sum_{k=1}^{N_p} c_k \,\tilde{r}_{hh}(n-k) + G\,\tilde{h}(-n), \qquad \text{all } n. \tag{11}$$

In the conventional LP case, we have a simplification because $h(n)$ is causal. For periodic signals we have to take into account terms in $\tilde{h}(n)$ for negative n.

For periodic signals, we can have a periodic correlation matching condition,

$$\tilde{r}_{hh}(n) = r_{xx}(n), \qquad 0 \le n \le N_p. \tag{12}$$

Substituting this into Eq. (11), we get a set of vector-matrix equations analogous to Eq. (1)

$$\boldsymbol{R}_{\mathbf{x}}\boldsymbol{a} = G\,\tilde{\boldsymbol{h}}, \tag{13}$$

where the elements of $\tilde{\boldsymbol{h}}$ are $\tilde{h}(0), \tilde{h}(-1), \ldots, \tilde{h}(-N_p)$. These equations are non-linear since $\tilde{\boldsymbol{h}}$ depends on $\boldsymbol{a}$. The $\boldsymbol{R}_{\mathbf{x}}$ matrix contains the (periodic) correlation for the observed data sequence.

If the observed correlation was not generated by an all-pole filter with the same number of terms as the analysis filter, then it may not be possible to match the aliased correlation values. It is shown in [1] that so-called singular cases can occur. An iterative procedure is used to find the solution, starting from the solution corresponding to the conventional LP analysis. Since the initial solution is known to have its roots inside the unit circle (minimum phase), El-Jaroudi and Makhoul suggest stopping the iteration short of convergence to avoid the singular case.

Consider fitting the 12'th order spectrum considered earlier. The input is periodic (32 sample period) and exact correlations are available. With a 12'th order fit, the DAP method eventually converges to the correct solution, while conventional LP misses the mark. When the fit is reduced to 8'th order, the DAP method gives reasonable matches at the harmonics of the spectrum but generates sharp peaks between the harmonics as illustrated in Fig. 4.

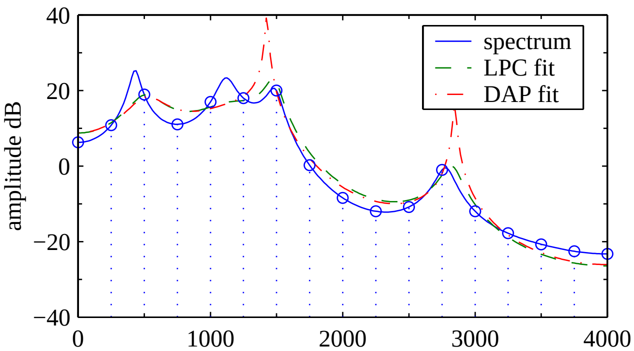

Fig. 4 Comparison of 8'th order all-pole fits to a 12'th order all-pole spectrum for a periodic input ($P = 32$). The solid line is the true spectral envelope, the dashed line is the conventional LP fit, and the dash-dot line is the fit using a Discrete All-Pole match.

2.2 Overdetermined approach for DAP

We can augments the DAP formulation to give an overdetermined set of equations. The error to be minimized is

$$\varepsilon_p^2 = \min_{\boldsymbol{a},G} \|\boldsymbol{W}^{1/2}(\boldsymbol{R}_{\mathbf{x}}\boldsymbol{a} - G\,\tilde{\boldsymbol{h}})\|^2, \tag{14}$$

where the matrix $\boldsymbol{R}_{\mathbf{x}}$ is rectangular (N_v+1 by N_p+1) and $\tilde{\boldsymbol{h}}$ is $[\tilde{h}(0), \tilde{h}(-1), \ldots, \tilde{h}(-N_v)]^T$.

The periodic correlation considered earlier will now be matched using an 8'th order DAP model using 12 equations. Fig. 5 shows that the effect of using more equations is to "tame" the DAP spectrum.

The concept of an overdetermined match is particularly useful if as in the example, the signal is generated by an all-pole filter of higher order than used in analysis. For systems where there is a good match in order, the overdetermined approach does not usually help and can sometimes make the match somewhat worse. Our experiments have shown

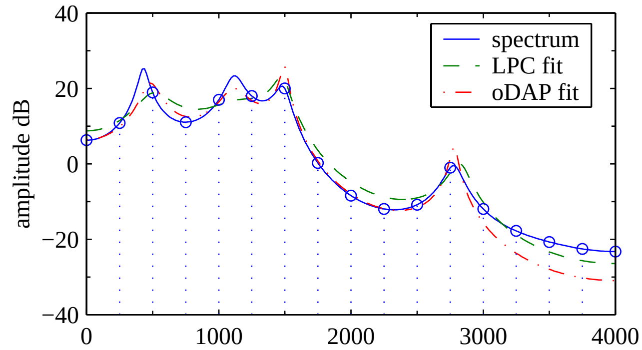

Fig. 5 Comparison of 8'th order all-pole fits to a 12'th order all-pole spectrum for a periodic input ($P = 32$). The solid line is the true spectral envelope, the dashed line is the conventional LP fit, and the dash-dot line is the fit using a Discrete All-Pole match minimizing an overdetermined equation criterion.

that choosing a small weight (much smaller than the values shown for the earlier example) for the extra equations still helps the case where the model order is too small, but does not change the solution much when the model order is adequate.

3 Mixed signals

The effective input to the speech synthesis filter for real speech signals is not perfectly periodic or perfectly random. Let a mixed signal be applied to the filter $H(z)$. The input signal is

$$\tilde{e}(n) = P \sum_{l=-\infty}^{\infty} \delta(n - lP + \theta) * h_L(n) + \eta(n) * h_H(n). \tag{15}$$

This signal is the sum of a (lowpass) filtered periodic impulse train, and a (highpass) filtered zero-mean white noise sequence (unit variance). The periodic part has a random time offset, $0 \le \theta < P-1$ and has been normalized to have a unit average energy per period. The two filters are chosen to be power complementary and include scaling to control the amount of periodicity as a function of frequency. At the output of the filter, the correlation of the sequence consists of two additive terms,

$$\tilde{r}_{hh}(k) = r_{hh}(k) + \sum_{\substack{l=-\infty \\ l\neq 0}}^{\infty} r_{hh}(k - lP) * r_L(k), \tag{16}$$

where $r_L(k)$ is the correlation for the filter applied to the periodic input. The effect of a periodic component becomes more significant for correlations which die off slowly and/or for short periods.

Extending the development that led to Eq. (3) to the mixed excitation case,

$$\tilde{r}_{hh}(n) = \sum_{k=1}^{N_p} c_k\, \tilde{r}_{hh}(n-k) + G\tilde{h}(-n), \tag{17}$$

where

$$\tilde{h}(-n) = h(-n) + \sum_{\substack{l=-\infty \\ l\neq 0}}^{\infty} h(-(n - lP)) * r_L(n). \tag{18}$$

This is consistent with the result for the purely periodic case when $r_L(k) = \delta(k)$ and for the purely random case when $r_L(k) = 0$.

This procedure assumes that the both the period and the periodicity factor are known. In speech coders, these parameters can be estimated as a byproduct of coding the pitch information.

Example

Consider a first-order filter with response

$$|H_L(\omega)|^2 = \frac{\rho_0 + \rho_\pi}{2} + \frac{\rho_0 - \rho_\pi}{2}\cos(\omega), \qquad (19)$$

where ρ_0 is the response at $\omega = 0$ and ρ_π is the response at $\omega = \pi$. The complementary filter has response $|H_H(\omega)|^2 = 1 - |H_L(\omega)|^2$. This parameterization allows the relative amounts of periodic and aperiodic excitation to be varied across frequency. These first-order filters have 3-term correlation functions.

For this example, the signal corresponds to a 10'th order spectrum. The mixed excitation signal is generated with $\rho_0 = 0.95$ and $\rho_\pi = 0.2$, with period varying from 30–60 samples. For analysis, we use a 240 sample Hamming window and use 10'th order analysis. Figure 6 shows the the spectral fits for conventional LP analysis. The fit is worst for the smallest period, since this corresponds to spectral lines which are far apart, or equivalently the correlations are most affected by aliasing.

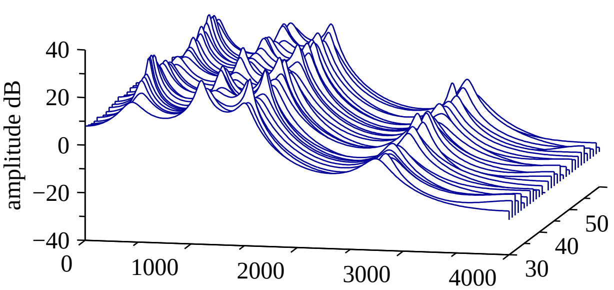

Fig. 6 Conventional 10'th order LP spectral fits for varying pitch period (240 sample Hamming window).

For comparison, the estimates in Fig. 7 were generated by the DAP algorithm run 5 iterations beyond the initial LP estimate. The DAP analysis gives significantly more consistent spectral estimates. For this example, the DAP algorithm was given information as to the periodicity factors and the period. Moderate mismatches of the periodicity factors do not seriously affect the fit, particularly since underestimation of the periodicity factor moves the solution towards that for conventional LP.

4 Summary and Conclusions

This paper has introduced methods for achieving better all-pole spectral estimates. For aperiodic (unvoiced) signals, an overdetermined equation approach provides estimates which are more consistent from frame to frame, particularly if the analysis order is too low. The overdetermined equation error approach also "tames" the discrete all-pole method for periodic signals.

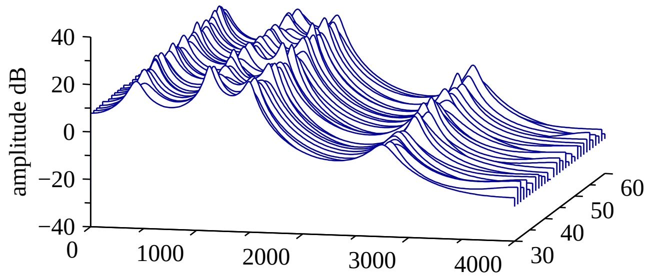

Fig. 7 DAP 10'th order spectral fits for varying pitch period (240 sample Hamming window).

A new approach to all-pole modelling of mixed signals better models this case, mitigating problems of formant / pitch interaction that are the bane of conventional LP analysis. In the examples, simple first order filters were used to control the periodicity across frequency. For some cases, more sophisticated separation may be warranted. For instance, in the case of MBE coding, the spectrum is considered to be periodic below a critical frequency and aperiodic above. This fits in neatly with the mixed excitation model for appropriately chosen lowpass and highpass filters.

More exact modelling of the interaction between formant and pitch can be achieved by estimating the pitch epoch offset (in our analysis, the results were obtained by averaging over all offsets), by modelling the frequency spreading effect of data windows, and by estimating the voicing mixture as a function of frequency.

References

[1] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Processing*, vol. 39, pp. 411–423, Feb. 1991.

[2] A. Das, A. V. Rao, and A. Gersho, "Enhanced multi-band excitation coding of speech at 2.4 kb/s with discrete all-pole spectral modeling," *Proc. IEEE Global Telecom. Conf.* (San Francisco), pp. 863–866, Dec. 1994.

[3] D. J. Molyneux, M. S. Ho, and B. M. G. Cheetham, "Robust application of discrete all-pole modeling to sinusoidal transform coding," *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing* (Istanbul), pp. 1455–1458, June 2000.

[4] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055–1096, Sept. 1982.

[5] L. B. Jackson, "AR modeling by least-squares fitting of windowed correlation data," *Proc. Third ASSP Workshop Spectral Est. and Mod.* (Boston), pp. 67–69, Nov. 1986.

[6] L. B. Jackson, *Digital Filters and Signal Processing.* Kluwer Academic, third ed., 1996.

[7] J. A. Cadzow, "ARMA spectral estimation: A model equation error procedure," *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing* (Denver, CO), pp. 598–602, Apr. 1980.