DISCRETE WEIGHTED MEAN SQUARE ALL-POLE MODELING

Davor Petrinovic
davor.petrinovic@fer.hr

Faculty of Electrical Engineering and Computing
University of Zagreb, Croatia

ABSTRACT

The paper presents a new method for all-pole model estimation based on minimization of the weighted mean square error in the sampled spectral domain. Due to discrete nature of the proposed distance measure, emphasis can be put on an arbitrary set of spectral samples what can greatly improve the model accuracy for periodic signals. Weighting can also be applied to improve the fitting in certain spectral regions according to any desired fidelity criterion. Iterative algorithm for determination of the optimal model is proposed and an exceptionally fast convergence rate is demonstrated. Accuracy of the estimation algorithm is verified on an example of a synthetic vowel for a broad range of pitch frequencies.

1. INTRODUCTION

In conventional LPC estimation techniques, coefficients of the all-pole model are determined by minimizing the Itakura-Saito (I-S) distance measure between the signal spectrum and the all-pole model spectral magnitude integrated across the whole spectrum. Indeed, such minimization yields exactly the desired result as long as the spectrum of the excitation signal is flat, i.e. single impulse or white noise signal. However this simplistic assumption does not hold for real speech signals, so the estimation accuracy can be severely limited, especially for periodic excitation signals (e.g. voiced speech segments). Since such signals have a discrete spectrum, frequency response of the vocal tract can be measured only at frequencies corresponding to the harmonics of the fundamental frequency. For all other frequencies, behavior of the system is unknown and must not be used in the all-pole estimation. Better estimation accuracy can be obtained by fitting the all-pole envelope only to the set of known frequency response samples of the system. The excitation signal is still assumed to be flat, but discrete, i.e. consisting of spectral lines.

The model that minimizes the average weighted mean square (WMSE) spectral distance in decibels between the signal spectrum samples and the samples of the all-pole model evaluated at the same set of frequency points is proposed in this paper. This technique utilizes summation in averaging of the estimation error, instead of integration as in the conventional LPC. It can be applied to an arbitrary set of frequency samples either harmonically related or not. Desire to use the WMSE spectral distortion measure in all-pole estimation has been present in the speech related field from its beginning, but the nonlinearity of the problem and nontractability of the solution were the main advocates against it. However, this paper will prove that the solution can be obtained in 'almost' closed form and that it is simple enough to be implemented in most of the speech applications. The benefit of the improved estimation accuracy offered by the method is equally important in speech analysis and coding, as well as in the feature extraction for speech recognition.

2. BACKGROUND

The interaction between fine spectral structure and the speech spectral envelope causes aliasing in the domain of correlation coefficients. Format frequencies of the estimated all-pole model derived from these coefficients are biased towards the pitch harmonics. Secondly, aliasing causes significant underestimation of formant bandwidths especially for high-pitched female voices. Different approaches had been tried out in attempt to solve the above problem such as: bandwidth expansion [1][2], new error criterion that are better suited to the statistical properties of the excitation signal [3], or different constrained versions of a linear prediction with certain undesired parts of the speech signal excluded from the correlation computation [4]. Some other approaches were based on smoothing of the speech spectrum by interpolation of some low order polynomial functions between harmonic peeks of the spectrum and calculating the LP model corresponding to the smoothed spectrum [5],[6].

Although all of the above methods give some improvement and some of them are even quite popular, the first attempt to solve the problem directly was the discrete all-pole modeling method DAP [7]. Discrete nature of the speech signal spectrum and the aliasing of

the correlation coefficients were taken into account during the all-pole estimation. The resulting model was the best fit to the spectrum lines in the sense of minimizing the discrete version of the I-S distance measure. The approach presented in this paper is similar to DAP but with two significant differences. Firstly, instead of I-S measure, the discrete WMSE spectral distance measure is used. Secondly, instead of modifying the predictor coefficients, the estimation is performed in the domain of line spectral frequencies (LSFs). These two differences have significant impact on the properties of the algorithm, as it will be shown.

3. WMSE ALL-POLE ESTIMATION

The proposed method is iterative and calculates new estimation of the all-pole model parameters based on the current model and the set of spectral samples that should be fitted by the new model in the least weighted mean square sense. Spectral samples $\mathbf{Y}=[y_1,y_2,...,y_N]^\mathsf{T}$ given in decibels correspond to the finite set of frequencies (lines) ω_i, i=1,2,...N. For voiced speech, these samples are equal to the harmonic peeks, while the frequencies, ω_i, match the integer multiples of the pitch frequency.

The initial all-pole model, $H(z)$, can be found by any of the conventional LP techniques and is given as:

$$H(z)=\frac{\sigma}{1+a_1z^{-1}+\ldots+a_pz^{-p}}=\frac{\sigma}{A(z)} \tag{1}$$

Magnitude response of the model $H(z)$ expressed in dB is given as:

$$P(\omega)=20\cdot\log_{10}\left(\left|H(e^{j\omega})\right|\right) \tag{2}$$

$$=20\cdot\log_{10}(\sigma)-10\cdot\log_{10}\left(\left|A(e^{j\omega})\right|^2\right) \tag{3}$$

$$=G-10\cdot\log_{10}\left(\begin{array}{l}\dfrac{R^2(\omega)}{2}(1-\cos(\omega))+\ldots\\ \ldots+\dfrac{Q^2(\omega)}{2}(1+\cos(\omega))\end{array}\right) \tag{4}$$

where $R(\omega)$ and $Q(\omega)$ are polynomials in the variable $\cos(\omega)$ of the order $p/2$ with real roots $[x_2, x_4, \ldots x_p]$ and $[x_1, x_3, \ldots x_{p-1}]$ respectively, i.e.:

$$R(\omega)=f_1(\omega,x_2,x_4,\ldots,x_p)=\prod_{i=1}^{p/2}2(\cos(\omega)-x_{2i}) \tag{5}$$

$$Q(\omega)=f_2(\omega,x_1,x_3,\ldots,x_{p-1})=\prod_{i=1}^{p/2}2(\cos(\omega)-x_{2i-1}) \tag{6}$$

The roots x_1 to x_p are equal to the cosine of the line spectrum frequencies corresponding to the model $H(z)$. The variable G gives gain of the model expressed in dB. Thus, through equations (4) to (6), the dB magnitude response $P(\omega,\mathbf{X})$ of the all-pole model is expressed as a function of ω and parameter vector $\mathbf{X}=[x_1, x_2, x_3,\ldots x_p, G]^\mathsf{T}$. For simplicity, the gain G will be treated as p+1st parameter denoted as x_{p+1} in the sequel. The WMSE spectral distance of the model to the given set of spectral samples can now be expressed as:

$$D=D(\mathbf{X},\mathbf{Y})=\sum_{i=1}^{N}w_i\cdot\left(y_i-P(\omega_i,\mathbf{X})\right)^2 \tag{7}$$

where w_i is the weight of the i^{th} spectral sample. The weights can be chosen according to any given fidelity criterion as it will be discussed latter.

Direct minimization of D with respect to $\mathbf{X}$ yields a set of nonlinear equations in x_1 to x_{p+1}. Therefore, the function $P(\omega,\mathbf{X})$ is expanded in the Taylor series around the initial vector $\mathbf{X}_0$ and approximated by only the first two terms of expansion, i.e.:

$$P(\omega,\mathbf{X})\cong P(\omega,\mathbf{X}_0)+\mathbf{S}(\omega,\mathbf{X}_0)\cdot\left(\mathbf{X}-\mathbf{X}_0\right) \tag{8}$$

$\mathbf{S}$ is the row vector comprised of first derivative functions of $P(\omega,\mathbf{X})$ with respect to components of $\mathbf{X}$ determined in the initial point $\mathbf{X}_0$. Since the distance D in (7) requires only the samples of the function $P(\omega,\mathbf{X})$ evaluated at frequencies ω_i, functions $P(\omega,\mathbf{X})$ and $P(\omega,\mathbf{X}_0)$ in (8) are replaced by column vectors $\mathbf{P}$ and $\mathbf{P}_0$ respectively. Row $\mathbf{S}(\omega,\mathbf{X}_0)$ becomes a sensitivity matrix denoted with $\mathbf{S}_0$, whose determination will be explained in the following section. For expressing the distance D in the matrix form, the weights w_1 to w_N are placed along the main diagonal of a diagonal matrix $\mathbf{W}$, that results with:

$$D=\left[\mathbf{Y}-\mathbf{P}\right]^\mathsf{T}\cdot\left[\mathbf{W}\right]\cdot\left[\mathbf{Y}-\mathbf{P}\right] \tag{9}$$

By assuming the equality in the expression (8), D can be rewritten in terms of variations of the model parameters $\Delta\mathbf{X}$ and variation of spectral values $\Delta\mathbf{Y}$:

$$D=\left[\Delta\mathbf{Y}-\mathbf{S}_0\cdot\Delta\mathbf{X}\right]^\mathsf{T}\cdot\left[\mathbf{W}\right]\cdot\left[\Delta\mathbf{Y}-\mathbf{S}_0\cdot\Delta\mathbf{X}\right] \tag{10}$$

$$\Delta\mathbf{X}=(\mathbf{X}-\mathbf{X}_0)\qquad\Delta\mathbf{Y}=(\mathbf{Y}-\mathbf{P}_0) \tag{11}$$

The problem of estimation can now be formulated as minimization of D with respect to $\Delta\mathbf{X}$. In other words, the goal is to modify the parameter vector in such a way that the new response $P(\omega,\mathbf{X})$ matches desired samples $\mathbf{Y}$ in the best possible way :

$$D_{\min}=\min_{\Delta\mathbf{X}}\left(D(\Delta\mathbf{X})\right) \tag{12}$$

Since D in (10) has the well-known quadratic form, the solution of (12) is straightforward, i.e. the parameter modification vector $\Delta\mathbf{X}$ must satisfy the following matrix equality:

$$\mathbf{\Phi}\cdot\Delta\mathbf{X}=\mathbf{\Psi} \tag{13}$$

where the system matrix $\mathbf{\Phi}$ and column vector $\mathbf{\Psi}$ are:

$$\mathbf{\Phi}=\mathbf{S}_0^\mathsf{T}\mathbf{W}\mathbf{S}_0\qquad\mathbf{\Psi}=\mathbf{S}_0^\mathsf{T}\mathbf{W}\cdot\Delta\mathbf{Y} \tag{14}$$

Once the solution of (13) has been found, the new parameter vector can be formed as:

$$\mathbf{X}_1=\mathbf{X}_0+\Delta\mathbf{X} \tag{15}$$

If the solution $\Delta\mathbf{X}$ is substituted back into the expression (10) for spectral distance D, it is easy to prove that D_{min} is given by:

$$D_{min} = \Delta\mathbf{Y}^{\mathsf{T}} \cdot \mathbf{W} \cdot \Delta\mathbf{Y} - \boldsymbol{\Psi}^{\mathsf{T}} \cdot \Delta\mathbf{X} \qquad (16)$$

The first term is the initial distance $D(\mathbf{X}_0,\mathbf{Y})$, while the second term is the reduction due to the correction vector $\Delta\mathbf{X}$. However, one should keep in mind that this is only an estimate of the distance, since it is based on the approximate model of $P(\omega,\mathbf{X})$ in (8). Therefore, to evaluate the actual distance of the new model $\mathbf{X}_1$, $P(\omega,\mathbf{X}_1)$ should be evaluated according to equations (4) to (6).

For non-negative weights, the $p\times p$ matrix $\boldsymbol{\Phi}$ is symmetric, positive semi-definite matrix, such that any efficient solution method can be applied. Numerical complexity of the solution is comparable to the covariance LP procedure. Since the minimization is reduced to the very well known least squares problem given in (13) and (14), all aspects concerning regularity of the matrix $\boldsymbol{\Phi}$ and uniqueness of the solution can also be applied here and will not be discussed.

Modification of the all-pole model is performed by modifying the LSF vector as in (15), so there is a potential risk that the new model might become unstable. However, this problem has been well studied in the LSF quantization literature and there are several safety techniques for ensuring the stability of H(z) in the LSF domain, that can also be applied in this algorithm after modification in (15).

4. CONVERGENCE OF THE SOLUTION

Since only the approximation of $P(\omega,\mathbf{X})$ was used in the minimization procedure, the new model isn't necessary the best solution, but nevertheless it is much better estimate then the initial one. If desired, the whole procedure can be repeated by expanding $P(\omega,\mathbf{X})$ around the new point $\mathbf{X}_1$, finding the new sensitivity matrix $\mathbf{S}_1$, new correction $\Delta\mathbf{X}$ and new model $\mathbf{X}_2$. Each iteration like this will yield better and better estimate.

The convergence rate depends on the accuracy of approximation of $P(\omega,\mathbf{X})$. Due to the favorable spectral sensitivity properties of the LSFs, the accuracy of (8) for small variations $\Delta\mathbf{X}$ is much better then for the other parametric representations of the all-pole model [8]. This constitutes one of significant advantages of the proposed method compared to [7] in which the modification is performed directly on the predictor coefficients. It will be shown on the example of a synthetic vowel that a single WMSE iteration of the proposed method performed on the initial autocorrelation LP model is sufficient to reduce the modeling error practically to zero, irrespectively of the pitch frequency.

5. SPECTRAL SENSITIVITY OF THE MODEL

Spectral sensitivity matrix $\mathbf{S}=\{s_{i,j}\}$ is necessary for computation of the matrix $\boldsymbol{\Phi}$ and the column vector $\boldsymbol{\Psi}$. The matrix has N rows for each of ω_i and p+1 columns corresponding to LSFs x_1, x_2, x_3,... x_p, and to the gain term $x_{p+1}=G$. The first derivatives of $P(\omega,\mathbf{X})$ with respect to components of $\mathbf{X}$ are readily available from the equations (4) to (6) as:

$$\frac{\partial P(\omega,\mathbf{X})}{\partial x_j} = \frac{(1-\cos\omega)R_j^2(\omega)}{C(\omega,j)}\ ,\ j=2,4,\ldots,p \qquad (17)$$

$$\frac{\partial P(\omega,\mathbf{X})}{\partial x_j} = \frac{(1+\cos\omega)Q_j^2(\omega)}{C(\omega,j)}\ ,\ j=1,3,\ldots,p-1 \qquad (18)$$

$$C(\omega,j) = \frac{\ln(10)}{40(\cos\omega - x_j)}\left| A\left(e^{j\omega}\right)\right|^2 \qquad (19)$$

$$\frac{\partial P(\omega,\mathbf{X})}{\partial x_{p+1}} = \frac{\partial P(\omega,\mathbf{X})}{\partial G} = 1 \qquad (20)$$

Functions $R_j(\omega)$ and $Q_j(\omega)$ are similar to $R(\omega)$ and $Q(\omega)$ in (5) and (6), but are missing one product term, i.e.:

$$R_j(\omega) = \prod_{\substack{i=1\\ 2i\neq j}}^{p/2} 2\left(\cos(\omega) - x_{2i}\right) \qquad (21)$$

$$Q_j(\omega) = \prod_{\substack{i=1\\ 2i-1\neq j}}^{p/2} 2\left(\cos(\omega) - x_{2i-1}\right) \qquad (22)$$

Finally the matrix elements $s_{i,j}$ are evaluated as:

$$s_{i,j} = \frac{\partial P(\omega_i,\mathbf{X})}{\partial x_j} \qquad (23)$$

Since all even columns have a lot of common factors of (21), significant complexity reduction can be obtained in computation. The same is also true for odd columns and the expression (22). Furthermore, the $\cos(\omega_i)$ term can be evaluated and stored only once for i=1 to N and then used in all of the above equations.

6. SPECTRAL WEIGHTING

Formulation of the proposed WMSE all-pole modeling includes spectral weights w_i. If all input spectral samples lay exactly on any given all-pole envelope of the order p, then the exact model with D=0 can be found in the minimization procedure for an arbitrary choice of w_i.>0, i=1 to N. However, for the real speech signals this is never true and the final model will only approximate the input samples in the WMSE sense. In this case weights can be used to put the emphasis on any given part of the spectrum or any given spectral sample according to some perceptual criterion, e.g. [7]. The weights for the new estimation iteration can even be derived from the coefficients of the current predictor estimate. For example, any of the LP derived weightings that are commonly used for quantization of the prediction residual in CELP coders, can be used as convenient choice. This way, not only the quantization of the LP residual but also the estimation of the all-pole model itself becomes consistent with the final distortion measure.

7. EXPERIMENTAL RESULTS

To evaluate the proposed WMSE all-pole modeling, an experiment was performed on a synthetic vowel '**a**' as in '**fa**ther'. The filter $H_a(z)$ was excited with an ideal synthetic periodic signal with linearly increasing pitch from 80 Hz to 300Hz and a total duration of 20 sec @ 8kHz. Spectral magnitudes of all excitation harmonics were exactly the same. The output of the filter was used as an input to the autocorrelation LP analysis with 200 samples Hamming window. The analysis was repeated for each 40 new time samples, resulting in the total number of 4000 frames. For each frame k, the estimated LP model $H_k(z)$ was compared to the template model $H_a(z)$, by computing the conventional SD distortion measure as:

$$SD(k) = \frac{20}{\sqrt{\pi}} \sqrt{\int_{\omega=0}^{\pi} \left(\log_{10} \left| \frac{H_k(e^{j\omega})}{H_a(e^{j\omega})} \right| \right)^2 d\omega} \qquad (24)$$

The LP model was then used as an initial all-pole model for 10 iterations of DAP estimation [7] and 2 iterations of the proposed WMSE estimation, both with unit weights. It was observed that DAP modeling has problems with convergence for default step sizes, so the step size reduction factor α=0.6 was applied as suggested in [7]. In the case of WMSE modeling, the optimal step size is computed automatically according to (13). For both techniques, the model mismatch was computed for each frame and iteration according to (24) and is plotted in Fig. 1. and 2. as a function of the pitch frequency. It can be observed that DAP estimation fails to resolve the template model even after 10 iterations, while for WMSE modeling the estimation error is bellow 0.35 dB after the first iteration and practically equal to zero after the second one.

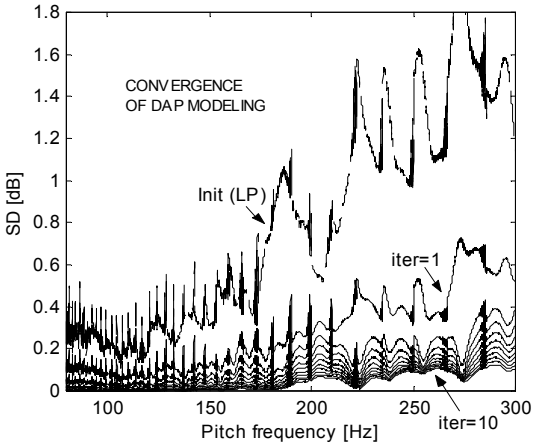

Figure 1. Estimation error of the DAP modeling

8. CONCLUSION

It has been demonstrated that WMSE distance measure in the sampled spectral domain can be successfully applied for the all-pole model estimation. The proposed technique resolves problems associated to the deficiencies of the conventional linear prediction methods and significantly improves the spectral match even in a single iteration. Interframe LSF quantization techniques can greatly benefit form WMSE estimation, since random fluctuations due to modeling errors are reduced. Residual signal quantization can also be improved since WMSE all-pole model performs better whitening of the spectral peeks then the conventional models. Furthermore, since the complexity of the method is comparable to the covariance LP method, it can be easily applied to any real-time system.

9. REFERENCES

[1] Tohkura, Y., Itakura, F., Hashimoto, S., "Spectral smoothing techniques in PARCOR speech analysis-synthesis", *IEEE Transaction on Acoustics, Speech And Signal Processing*, vol. ASSP-26, no. 6, pp. 587-596, December 1978

[2] Viswanathan, R., Makhoul, J., "Quantization properties of transmission parameters in linear predictive systems", *IEEE Transaction on Acoustics, Speech And Signal Processing*, vol. ASSP-23, pp. 309-321, June 1975

[3] Lee, C.H., "On robust linear prediction of speech", *IEEE Transaction on Acoustics, Speech And Signal Processing*, vol. ASSP-36, no. 5, pp. 642-650, 1988

[4] Miyoshi, Y., Yamato, K., Mizoguchi, R., Yanagida, M., Kakusho, O., "Analysis of speech signals of short pitch period by a sample-selective linear prediction", *IEEE Transaction on Acoustics, Speech And Signal Processing*, vol. ASSP-35, no. 9, pp. 1233-1240, 1987

[5] Hermansky, H. "Spectral envelope sampling and interpolation in linear predictive analysis of speech", Proc. IEEE ICASSP,1984, pp. 2.2.1-2.2.4

[6] McAulay, R.J., Quatieri, T.F., "Sinusoidal Coding" in *Speech Coding and Synthesis*, ed. Kleijn, W.B., Paliwal, K.K., Elsevier, 1995, pp. 121-173

[7] El-Jaroudi, A., Makhoul, J., "Discrete All-Pole Modeling", *IEEE Transaction on Signal Processing*, vol. 39, no. 2, pp. 411-423, Feb. 1991

[8] Gardner, W.R., Rao, B.D., "Theoretical analysis of the high-rate vector quantization of LPC parameters", *IEEE Transaction on Speech And Audio Processing*, vol. 3, no. 5, pp. 367-381, Sep. 1995

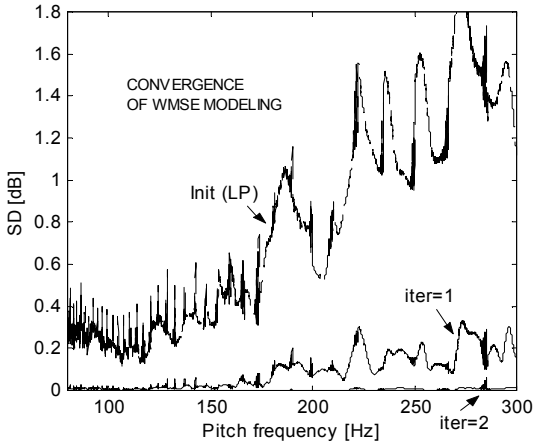

Figure 2. Estimation error of the WMSE modeling